Hydroclimatology of Extreme Precipitation and Floods Originating from the North Atlantic Ocean

Jennifer Nakamura

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

Hydroclimatology of Extreme Precipitation and Floods Originating from the North Atlantic Ocean

Jennifer Nakamura

This study explores seasonal patterns and structures of moisture transport pathways from the North Atlantic Ocean and the Gulf of Mexico that lead to extreme large-scale precipitation and floods over land. Storm tracks, such as the tropical cyclone tracks in the Northern Atlantic Ocean, are an example of moisture transport pathways. In the first part, North Atlantic cyclone tracks are clustered by the moments to identify common traits in genesis locations, track shapes, intensities, life spans, landfalls, seasonal patterns, and trends. The clustering results of part one show the dynamical behavior differences of tropical cyclones born in different parts of the basin. Drawing on these conclusions, in the second part, statistical track segment model is developed for simulation of tracks to improve reliability of tropical cyclone risk probabilities. Moisture transport pathways from the North Atlantic Ocean are also explored though the specific regional flood dynamics of the U.S. Midwest and the United Kingdom in part three of the dissertation.

Part I. Classifying North Atlantic Tropical Cyclones Tracks by Mass Moments.

A new method for classifying tropical cyclones or similar features is introduced. The cyclone track is considered as an open spatial curve, with the wind speed or power information along the curve considered as a mass attribute. The first and second moments of the resulting object are computed and then used to classify the historical tracks using

standard clustering algorithms. Mass moments allow the whole track shape, length and location to be incorporated into the clustering methodology. Tropical cyclones in the North Atlantic basin are clustered with K-means by mass moments producing an optimum of six clusters with differing genesis locations, track shapes, intensities, life spans, landfalls, seasonality, and trends. Even variables that are not directly clustered show distinct separation between clusters. A trend analysis confirms recent conclusions of increasing tropical cyclones in the basin over the past two decades. However, the trends vary across clusters.

Part II: Tropical cyclone Intensity and Track Simulator (HITS) with Atlantic Ocean Applications for Risk Assessment.

A nonparametric stochastic model is developed and tested for the simulation of tropical cyclone tracks. Tropical cyclone tracks demonstrate continuity and memory over many time and space steps. Clusters of tracks can be coherent, and the separation between clusters may be marked by geographical locations where groups of tracks diverge due to the physics of the underlying process. Consequently, their evolution may be non-Markovian. Markovian simulation models, as often used, may produce tracks that potentially diverge or lose memory quicker than nature. This is addressed here through a model that simulates tracks by randomly sampling track segments of varying length, selected from historical tracks. For performance evaluation, a spatial grid is imposed on the domain of interest. For each grid box, long-term tropical cyclone risk is assessed through the annual probability distributions of the number of storm hours, landfalls, winds, and other statistics. Total storm length is determined at birth by local distribution, and movement to other tropical cyclone segments by distance to neighbor tracks,

comparative vector, and age of track. An assessment of the performance for tropical cyclone track simulation and potential directions for the improvement and use of such model are discussed.

Part III: Dynamical Structure of Extreme Floods in the U.S. Midwest and the United Kingdom.

Twenty extreme spring floods that occurred in the Ohio Basin between 1901 and 2008, identified from daily river discharge data, are investigated and compared to the April 2011 Ohio River flood event. Composites of synoptic fields for the flood events show that all these floods are associated with a similar pattern of sustained advection of low-level moisture and warm air from the tropical Atlantic Ocean and the Gulf of Mexico. The typical flow conditions are governed by an anomalous semi-stationary ridge situated east of the US East Coast, which steers the moisture and converges it into the Ohio Valley. Significantly, the moisture path common to all the 20 cases studied here as well as the case of April 2011 is distinctly different from the normal path of Atlantic moisture during spring, which occurs further west. It is shown further that the Ohio basin moisture convergence responsible for the floods is caused primarily by the atmospheric circulation anomaly advecting the climatological mean moisture field. Transport and related convergence due to the covariance between moisture anomalies and circulation anomalies are of secondary but non-negligible importance. The importance of atmospheric circulation anomalies to floods is confirmed by conducting a similar analysis for a series of winter floods on the River Eden in northwest England.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I did not do any of this research alone. Upmanu Lall and Yochanan Kushnir have lent me constant support, productive ideas, and a vast knowledge of climate and statistics making this dissertation possible.

Part I

Part I added the hurricane expertise of Suzana J. Camargo. Reviewers added constructive comments that improved the work. Analysis was made possible by the HURDAT data set. I acknowledge the many hurricane hunters who have flown into the storms collecting data to contribute to HURDAT including the 11 lost crew of the U.S. Navy P2V Neptune weather reconnaissance airplane Snowcloud Five. An award from the Columbia University Office of Research Initiatives funded the study.

Part II

Part II both started and ended with Balaji Rajagopalan's positive contributions. Analysis was made possible by Suzana's Matlab converted HURDAT data set. The work was supported by NSF grant AGS-1003417 and NOAA Grants NA10OAR4310212 and NA10OAR4310159.

Part III

Part III added the climate expertise of both Andrew W. Robertson and Richard Seager, along with Richard's editing and organizing talents. I would like to thank two anonymous reviewers for their constructive advice and guidance. Analysis was made possible by 20th century reanalysis and the USGS HCDN streamflow data set. The work was supported by NOAA grants NA10OAR4310159, NA10OAR4320137 and NA10OAR4310137 (Global Decadal Hydroclimate Variability and Change).

This page intentionally left blank

Dedicated to my entire family, dear friends, coworkers, colleagues,

and to anyone reading this sentence right now.

Thank you!

This page intentionally left blank

Chapter 1

Introduction

1.1 Motivation

Floods account for a large fraction of losses due to natural hazards. It is recognized that, especially for larger river basins, a link to an oceanic moisture source is likely necessary for generating the extreme precipitation associated with a large regional flood (Hirschboeck 1991). Prior moisture transport into the basin from the same oceanic source may have saturated soils leading to enhanced flood potential. For extreme floods, large-scale meteorological conditions associated with organized transport of moisture from an ocean to a continental region are likely important. Grouping floods and moisture transports on the seasonal scale over many years gives a hydroclimatology perspective of changes in the frequency, magnitude, duration and location of floods as they interact with large-scale atmospheric circulation patterns.

Storm tracks, such as the tropical cyclone tracks in the Northern Atlantic Ocean, are an example of moisture transport pathways from an oceanic source. Although significant flooding is caused by storm surges, rainfall is also an important contributor to floods. Shepherd et al. (2007) found that even though they are a rare event, tropical cyclones

accounted for a mean of 12.8% of the cumulative rainfall for the hurricane season for southeast U.S. (SE) coastline basins and the major tropical cyclones are apparent in extreme daily events. Smith et al. (2010), in a study of the Delaware basin, found that the August 1955 flood of Hurricane Diane is the standing peak flow record for 40% of the basin.

Other than hurricane season in the SE, spring is the season with the greatest flooding potential in most of the United States (Hirschboeck 1991). The southern part of the U.S. is warming in spring, while the north is cool and upper level wind energy is still large as in winter. The frozen ground can decrease infiltration and snowmelt adds to the rainfall totals. The Ohio River is the largest tributary, by volume, of the Mississippi River. This study chooses large (over 10,000 km^2) sub-basins in the Ohio basin that are not effected by dams or other human regulation in order to connect spring floods to patterns of moisture transport from the oceans. These transports are also compared to similar wintertime floods in the U.K. with moisture also originating from the North Atlantic Ocean. Knippertz and Wernli (2010) and Knippertz et al. (2013), in a climatology of tropical moisture exports (TME), refers to the U.S. moisture transports as the Great Plains tropical moisture exports (TME) with a maximum in spring and summer, and the U.K. moisture transports as the Gulf Stream TME with a maximum in winter and autumn.

1.2 Problem Statement and Approach

The following research into the hydroclimatology of extreme precipitation and floods originating from the North Atlantic Ocean seeks to answer the questions:

- Do North Atlantic tropical cyclone tracks show similarities that would allow improved forecasting of their attributes?
- Can these track similarities lead to better risk assessment probabilities?
- What other patterns of oceanic moisture transport lead to large-scale floods?

Standard K-means is a clustering method that minimizes the distance from the data points to the cluster. Tropical cyclone tracks vary in length making standard K-means clustering not possible. The approach of Part I is to reduce the length, latitude, and longitude information of North Atlantic tropical cyclone tracks through moments into a standardized fixed length matrix allowing clustering. The characteristics of genesis locations, track shapes, intensities, life spans, landfalls, seasonality, and trends can be examined in each grouping. Dimension reduction is a common method used in clustering and data mining. In this case, the novel use of moments allows contraction of an entire tropical cyclone track into a few relevant attributes.

Tropical cyclones are relatively rare in the North Atlantic, with about ten each season. The chance of impacting land is even lower and lack of data leads to uncertainty in the landfall probabilities. These two sources of zero yearly landfall, overall tropical cyclone rarity in the basin and steering away from land, give historical distributions are zero-inflated and heavily positively skewed to lower values. The approach of Part II is to create a statistical model, based only on the North Atlantic best track dataset (HURDAT) data and knowledge of the track similarities, to create many more plausible tracks and more accurate probabilities. Statistical models that resample historical data are used in a

variety of applications to find the unknown distribution of a variable, such as in the Monte Carlo method.

Next, using a threshold of a 10-year flood event in large basins (over 10,000 km^2) the pattern of moisture transport to the middle of the U.S. is explored and compared to those in the U.K. The approach of Part III is to composite the meteorological fields of the days leading up to 20 extreme flood events in the selected basins looking for a coherent signal in all events. Time is set relative to the event, i.e. 1 day prior, 2 days prior, etc. up to nine days prior. Compositing is a simple, yet effective way to find a common response in a set of data that contains noise from other influences. Examining the prior days allows a view of the common prior steps to the event. Decomposition of the anomalies into climatological and anomalous components leads to understanding the driving factors of these floods.

1.3 Overview of the dissertation

This dissertation is organized in four parts and is visually represented by Fig 1.1.

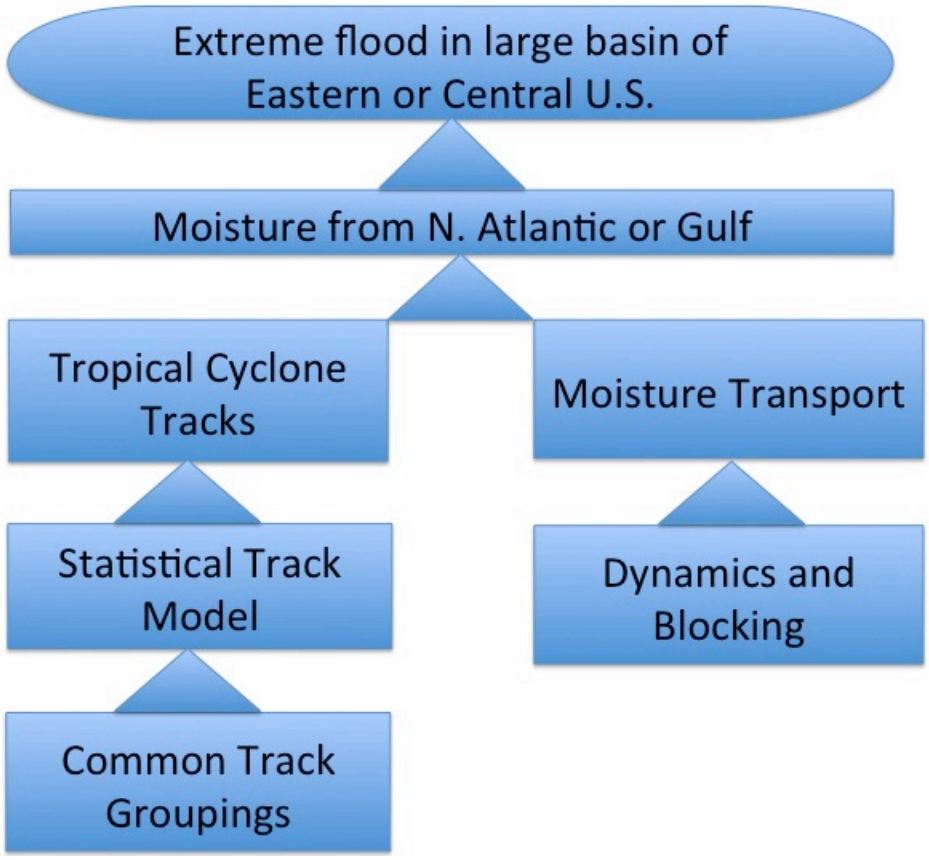
In Part I, North Atlantic tropical cyclone tracks are converted by moments into a small number of attributes, which are then K-means clustered finding common track groupings. Genesis locations, track shapes, intensities, life spans, landfalls, seasonality, and trends are analyzed for each cluster. This leads to Part II, where a nonparametric stochastic model is developed and tested for the simulation of North Atlantic tropical cyclone tracks that simulates tracks by randomly sampling track segments of varying length, selected from historical tracks. For performance evaluation, a spatial grid is imposed on the domain of interest. For each grid box, long-term tropical cyclone risk is assessed through

the annual probability distributions of the number of storm hours, landfalls, winds, and other statistics. An assessment of the performance for tropical cyclone track simulation and potential directions for the improvement and use of such model are discussed.

In Part III, oceanic tropical moisture transports are explored by compositing 20 spring 10-year flood events in the Ohio basin and compared to an event in April 2011 exploring the topics of dynamics and blocking. Composites of precipitation, 700 mb geopotential heights, 900 mb specific humidity, and climatological and anomalous moisture flux and moisture convergence are presented. The moisture flux anomaly is decomposed into climatological circulation transporting the moisture anomaly, circulation anomaly transporting the climatological humidity, and the circulation anomaly transporting the moisture anomaly to find the dominant term. A dipole index is computed with 700 mb height centers on days prior to the flood. North Atlantic moisture transports into the U.S. are then compared with those in wintertime U.K. through moisture fluxes and the decomposition of terms.

Parts I, II, and III are examples of the moisture transports from the N. Atlantic or Gulf of Mexico that cause extreme flooding in large basins of the Eastern or Central U.S. as well as in the U.K. Finally, Part IV of this dissertation discusses the contributions of this study, and future work.

Figure 1.1: Diagram of dissertation organization.



This page intentionally left blank

Part I

Classifying North Atlantic Tropical Cyclones Tracks by Mass Moments.

This page intentionally left blank

A new method for classifying tropical cyclones or similar features is introduced. The cyclone track is considered as an open spatial curve, with the wind speed or power information along the curve considered as a mass attribute. The first and second moments of the resulting object are computed and then used to classify the historical tracks using standard clustering algorithms. Mass moments allow the whole track shape, length and location to be incorporated into the clustering methodology. Tropical cyclones in the North Atlantic basin are clustered with K-means by mass moments producing an optimum of six clusters with differing genesis locations, track shapes, intensities, life spans, landfalls, seasonality, and trends. Even variables that are not directly clustered show distinct separation between clusters. A trend analysis confirms recent conclusions of increasing tropical cyclones in the basin over the past two decades. However, the trends vary across clusters.

Citation: Nakamura, J., U. Lall, Y. Kushnir and S. J. Camargo, 2009: Classifying North Atlantic Tropical Cyclone Tracks by Mass Moments. *Journal of Climate*, **22**(20): 5481-5494.

This page intentionally left blank

Chapter 2

Introduction to Part I

Tropical cyclones lead to major natural disasters in the regions of landfall with devastating storm surges, flooding, and high winds. Named storms (NS) include all tropical cyclones reaching a maximum sustained wind speed of at least 18 m s^{-1} or 35 kts (Neumann, et al. 1993). Their impact along the Atlantic Coast of North and Central America is often catastrophic to life, ecology, property, wetlands and coastal estuaries (e.g., in 1995 total mainland U.S. tropical cyclone damages averaged the order of $5 billion – Pielke and Landsea, 1998). The severity and frequency of NS is consequently of great interest for disaster planning and mitigation. Determining the spatial characteristics of tropical cyclone frequency and intensity is of considerable importance for any study of past, present, and future tropical cyclone impact or model validation. Clustering provides a way to assess congruence in spatial characteristics such as track shape, genesis location, intensity, lifespan, seasonality, and landfall. The causal factors associated with each cluster as well as the ability of numerical simulation models to simulate the frequency of each cluster, conditional on specified boundary conditions could then be assessed. Macro

level statistics of the conditional dependence of each cluster on antecedent conditions could also be assessed.

To a first approximation, tropical cyclones move in the direction that the mean tropospheric winds (over the depth of the storm) steer them. In the Northern Atlantic, the northeasterly trade winds move the storms westward from the African coast. The prevailing flow around the subtropical high curves them northward approaching the North American coast, and eastward in the middle latitudes. Elsner and Kara (1999) call this track shape the parabolic sweep. The position and strength of the subtropical high, the extra-tropical circulation, and the genesis location of the tropical cyclone vary, allowing variation in track shape and length. Genesis location can be linked to the seasonality, sea surface temperatures, wind shear, and position of the initial disturbance (Gray, 1968, 1979; Henderson-Sellers et al., 1998). Both maximum wind speed (or intensity) and lifespan are linked to the genesis location and track shape as some locations/curvatures provide a longer time to intensify before encountering land or colder water. Landfall and the intensity of the storm at landfall are also associated with genesis location and track shape (Camargo et al., 2007).

Cluster analysis provides a way to objectively classify storms in a given ocean basin into subcategories depending on geographical properties of the storms (e.g., genesis, track location, and shape etc.) Such classification can become useful for building predictive understanding on climate time scales. The K-means method (MacQueen, 1967) is a common clustering method that has been used both with tropical (Elsner and Liu, 2003; Elsner 2003) and extratropical cyclone tracks (Blender et al., 1997). The method is typically applied to vector data on select attributes, and seeks to find k-separations of the

data such that the inter-cluster variation is maximized, relative to the centroid of each cluster in the space formed by the vectors submitted. The application of this method to tropical cyclone tracks (potentially specified as vectors defined by the x and y coordinates of the track) also faces the challenge that the tracks, and hence the data vectors with track coordinates are of unequal lengths. The studies referenced above addressed this problem by using two points in the track (positions of maximum and final tropical cyclone intensities), or dimensional vectors (centroids of the cyclone trajectories every six-hours for the first three days) respectively.

In the present study the entire track shape and length are taken into account by using the mass moments of the open curve that defines a full storm track. The first two moments are used and are defined and illustrated below. The first mass moment is simply the location of the x and y coordinates of the centroid of an object (in this case an open curve, and the centroid lies in the area interior to the curve, but not on the curve – see star of Fig. 6.1). The second moment is the variance (x, y, and xy directions). The second moment is illustrated by an ellipse in Fig. 6.1. Mathematically, the two moments are expressed as follows:

First Moment:

$$\boldsymbol{M1} = \frac{1}{A}\int w(\boldsymbol{r})\boldsymbol{r}\,dx\,dy = \frac{1}{\sum_{i=1}^{n} w(r_i)}\sum_{i=1}^{n} w(\boldsymbol{r}_i)\boldsymbol{r}_i \qquad (1)$$

Where **r** is the coordinate vector (x,y) for a point on the track, w(**r**) is a weight associated with that location (e.g., for the case of tropical cyclone a measure of wind /storm

intensity), the integral is taken over the open curve that defines the track, A is an normalization constant that reflects the total intensity over the track, and the sum on the right hand side of the equation defines the discrete approximation to the integral over the track, using the n recorded locations of the track over the lifetime of the storm.

Second Moment:

$$\boldsymbol{M2} = \frac{1}{A}\int w(\boldsymbol{r})\{\boldsymbol{r} - \boldsymbol{M1}\}^2\, dxdy = \frac{1}{\sum_{i=1}^{n} w(r_i)}\sum_{i=1}^{n} w(\boldsymbol{r}_i)\{\boldsymbol{r}_i - \boldsymbol{M1}_i\}^2 \quad (2)$$

M1, r, w(**r**), A, and n are defined as in the first moment. In matrix format, the variance is the covariance matrix between the scalar components (x and y) of r giving three directions to the second moment of x, y, and xy.

These five numbers (two centroid and three covariance) then constitute the summary of the track information that is to be used to identify track clusters. The first moment establishes the location of effective center of gravity of the storm track, while the second moment provides a measure of the shape of the storm. The classical covariance measure is usually explained as a measure of the orientation and length of the principal axes of an ellipse that describes the scatter of data in a plane. A similar interpretation applies here, with the qualification that the data of interest is not a scatter plot, but a section of a curve that is being approximated as an ellipse. Thus, if two tracks were perfect straight lines but with different lengths, then the second moment matrix would identify the same orientation for both of them, but a different spread along that direction. If the first track is a straight line and the other part of an ellipse, then from the covariance ellipse we

would see that the length of the second principal axes is zero indicating that there is no variation in that direction, while it would be non-zero for the curve, the magnitude increasing as the relative curvature increases. Further, if both tracks are curves, but one is convex and the other concave with respect to a particular reference frame then the signs of the cross-covariance terms in the second moment matrix are opposite, and one can recognize this condition. Thus, using the first two moments of a storm track one can get a measure of its central location, length, orientation and also curvature. This allows one to get an approximation to most of the typical tracks that we observe. Complex features, such as tracks that curve back on to themselves or are concave in some sections and convex in others will require recourse to higher moments.

The analysis of the track's moments is performed here not during the evolution of a track, but post-mortem taking into account the whole life cycle. Hence, this approach will not be useful to describe an evolving storm. Rather, it is intended for an analysis of historical data, especially for the identification of trends, and associated causal factors for particular types of storms.

The data set used is described in the next section along with some aspects of the clustering methodology. Chapter 4 presents associated results for N. Atlantic tropical storms, including a discussion of the attributes of the clusters in terms of spatial characteristics such as track shape, genesis location, intensity, lifespan, seasonality, and landfall. Chapter 5 compares our results with those from a selection of other methods.

This page intentionally left blank

Chapter 3

Data

Atlantic basin historical hurricane track data (HURDAT), with information on storm position (latitude/longitude) every six-hours available from the National Hurricane Center (NHC) was used. Only NS from 1948-2006 were retained for our analysis. Due to routine aircraft reconnaissance missions into tropical cyclones beginning in 1944, details on the position of the tropical cyclone structure have been available. This has lead to greater accuracy in the six-hour position data.

3.1 Clustering Methodology

A variety of methods for the identification of clusters from vector data are available. Here, we use the K-means method with a vector of five attributes per track - two for longitudinal and latitudinal centroid and three for the variance (longitudinal, latitudinal, and the diagonal).

The variance components contain much larger values than the centroid components. In order to give the centroid and variance equal weight in the cluster analysis the variables are standardized (to mean zero and standard deviation 1) and each centroid column is multiplied by 0.5/2, and each variance column by 0.5/3, so that the centroid and the

covariance are treated roughly on par as to their importance for clustering. Other weights that emphasize one or the other attribute could indeed be chosen.

Moments of the N. Atlantic cyclone tracks were estimated using wind velocity (or power) recorded along the track as a weight and also using uniform weights (w(**r**) of ones in 1 and 2). Each of the resulting data was then used in K-means cluster analysis. Results are presented for the uniform weight case. Separation between the clusters was distinct and clusters cohesive as tracks were grouped together in geographical region. When wind velocity weights are used tracks farther apart are added to the cluster.

The K-means clustering algorithm partitions the data into k-clusters, with cluster centroids denoted by μ_i, and coordinates as $x_{i,j}$, such that the variance across clusters defined below is maximized. J is the index over all points in cluster i (i = 1…k).

$$Var = \sum_{i=1}^{k} \sum_{j=1}^{l} \left(x_{i,j} - \mu_i \right)^2 \quad (3)$$

The K-means cluster analysis package available in Matlab 7.3 considers multiple runs with random seeding of clusters. The optimal cluster number j is determined by the maximum mean and minimum number of negative "silhouette" values. A silhouette is both a measure of how cohesive each cluster is and how well the clusters are separated. For i total points a silhouette (S_i) is defined as

$$S_i = \frac{\min\left(b_i\right) - a_i}{\max\left[a_i, \min\left(b_i\right)\right]} \quad (4)$$

where a_i is the average distance from the ith point to the other points within the cluster and b_i is the average distance from the ith point to points in another cluster (Kaufman and Rousseeuw, 1990). Silhouette values range from -1 to 1. Clusters with a high mean silhouette value are cohesive and negative silhouette values are possible misclassified points. Figure 6.2 shows the mean silhouette values (top) and number of negative silhouette values (bottom) for a selected run with the N. Atlantic tropical clone data. Cluster number of two through nine run along the x-axis. Note that the mean silhouette value for the six clusters is a maximum in the top figure and the number of negative silhouette values a minimum in the bottom figure indicating the best choice of cluster number is six.

This page intentionally left blank

Chapter 4

Cluster Analysis Results

4.1 Centroid and Variance

The actual output of the K-means clustering is groups of centroid locations and directional variance (Fig. 6.3). Clear separation of the clusters can be seen in the groupings of the centroids and slope and size of the variance ellipses. Mean centroid value is marked with a dark x, and mean variance ellipse with a dark line. Table 6.1 presents the mean centroid location and directional variance for each cluster (1-6). Cluster 1 is centered just west of mid-basin with rounded variance ellipses (mean variance x approximately equal to mean variance y) and a slightly negative tilt (negative mean variance xy). The mean location of cluster 2 is slightly higher than that of cluster 1 near the eastern coast of the United States. The variance ellipses are elongated in x-axis with a positive tilt. Cluster 3 sits mid-basin with rounded ellipses and nearly zero tilt. The farthest southern cluster is 4 and the ellipses are stretched along the x-axis with a slightly negative tilt. The farthest north of all the clusters is 5. It also has the largest ellipses, elongated in the x-axis, and heavily positively tilted. Cluster 6 is centered in the Gulf and like cluster 3 is rounded with near zero tilt. It has the smallest variance ellipses.

4.2 Genesis location and Track Shape

Although the track location and shape are taken into account with the centroid and variance, the first reported location or genesis and actual track are not directly used in the entire analysis. Still, clear groupings can be seen in Fig. 6.4 with the genesis location with a circle and the track as a line. Table 6.2 presents the mean genesis location in each cluster along with the range of those first positions. Cluster 1 genesis locations fall roughly within the main development region of Goldenberg et al., (2001). These are Cape Verde hurricanes (Atlantic basin hurricanes that form close to the Cape Verde Islands (<1000 km) and become tropical cyclones before reaching the Caribbean) moving in a classic parabola shape as described by Elsner and Kara, (1999). They are the farthest east forming tropical cyclones of all the clusters and have the smallest genesis range in latitude. The farthest north forming tropical cyclones are in cluster 2. They have a diffuse genesis range in both latitude and longitude and move along the Eastern coast of the United States. The largest genesis range in latitude is found in cluster 3 and they follow a flattened parabola shape. The genesis location of cluster 4 resembles cluster 1 with a small genesis range in latitude in the Cape Verde region. What differentiates these tropical cyclones is the nearly straight track shape: they rarely venture above 25N. The most diffuse in genesis location is cluster 5 with the largest genesis range in longitude. The tropical cyclones in cluster 5 are not full parabolas, but rather partial "hooks" or semi-parabolas. Cluster 6 genesis locations are confined to the Gulf and Caribbean with the smallest genesis range in longitude. These tropical cyclones also tend to stay in the Gulf.

4.3 Intensity and lifespan

Lifespan of a storm usually affects the size of the variance ellipse; i.e. a longer track equals a larger ellipse unless the tropical cyclone is very slow moving. The next three figures that explore the maximum wind speed (Fig. 6.5), lifespan (Fig. 6.6), and maximum intensity (Fig. 6.7) are box plots. These illustrate the 25^{th} and 75^{th} percentile (upper and lower bounds of the box), the mean (star), the median (bar in middle), the bounds (dashed line), and outliers (plus signs) of the distribution of tropical cyclones in each cluster and as a whole.

Maximum wind speed is categorized by the Saffir-Simpson scale into tropical storm (TS) 35-64 kts, category 1 of 65-82 kts, 2 of 83-95 kts, 3 of 96-113 kts, 4 of 114-135 kts, and 5 of greater than 135 kts. Figure 6.5 shows the maximum wind speed in knots of the six clusters and all tropical cyclones. Table 6.3 presents the mean and median categories of the maximum wind speed. Cluster 1 is considerably stronger than the other clusters and total, with cluster 5 coming in second. The clusters 2 and 3 are only slightly weaker than the total mean keeping within the same category. Clusters 6 and 4 are weaker than the total with the median maximum wind speed falling in the TS category. The strongest cluster, 1, shows a negative skew distribution as compared to normal with the median in a higher category than the mean (majority of storms weaker than the mean) and the weakest clusters, 4 and 6, shows a positive skew with the median in a lower category than the mean (majority of storms stronger than the mean).

Lifespan was converted from 6-hr periods to days for easier viewing in Fig. 6.6. If ranked from longest to shortest it follows the same pattern as the maximum wind speed

except for the last two, which are nearly equal. The three longest clusters, 1, 5, and 3, are longer than the total and the three shortest, 2, 4, and 6 are shorter. The longer the life span the longer the time the tropical cyclone has to intensify as long as the conditions stay favorable (i.e. warm waters and low shear). Life spans were all negatively skewed with the majority of the tropical cyclones lasting less time than the mean.

The power dissipation index (PDI – Emanuel, 2005) is a hybrid of maximum wind speed and lifespan and is used here as a measure of integrated intensity. Emanuel (2007) found the index to covary with low-level vorticity and vertical wind shear and correlate highly with sea surface temperature. PDI is defined as the simplified power dissipation index as in Emanuel, (2005)

$$PDI \equiv \int_{1}^{n} V^3 dt \qquad (5)$$

where n is number of time steps with dt in seconds and V is the wind velocity in ms^{-1} giving units of PDI in m^3/s^2. The large index values are multiplied by $1*10^{-11}$ for plotting ease. Not surprisingly the combination of wind speed and lifespan shows cluster 1 to be the most intense, followed by cluster 5. These are the only two that are stronger than the average of all. However, the distributions of clusters 2-6 show quite a few outliers indicating that intense storms are possible in these clusters, but more of a rare event than for cluster 1. All the distributions are negatively skewed just like the life spans indicating that the majority of tropical cyclones are less intense than the mean.

4.4 Seasonality

If following the mariner's poem (Inwards, 1898),

June too soon.

July stand by.

August look out you must.

September remember.

October all over.

one would conclude the North Atlantic hurricane season to be four months long; July to October. Indeed, looking at a box plot of storm month (Fig. 6.8) and not counting outliers, three of the clusters (1, 3, and 6) along with all storms show a four-month season of July to October. However, two clusters have a six month season (4 and 5) of June to November and cluster 2 has an eight month season of April to November. Outliers even show a tropical cyclone in February. Note, due to the discrete nature of the month sometimes the 25^{th} or the 75^{th} percentile and the median are the same so for this figure the median is marked with a circle. The cluster means all fall in August (8) and September (9) with two medians in August (1 and 2), and the rest in September. September is to remember. Three clusters have a negative skew with more tropical cyclones occurring earlier than the mean (1, 2, and 3), and other three along with the total have a positive skew with more occurring later than the mean.

4.5 Landfall

Landfall was inferred by applying a one degree land-sea mask and counting the number of storms that crossed from ocean to land from 5N to 45N and 100W to 50W. Table 6.4 presents the cluster number, number of tropical cyclones in each cluster, number of landfalls, and the percentage of landfall. Not only does cluster 6 have the most number of storms, it also has the most landfalls. Tropical cyclones in that cluster do not have far to travel before they hit land. The most intense tropical cyclones in cluster 1 have a slightly higher percentage of hitting land than all storms; however, the number of storms in that cluster is fairly low. The second most intense storms in cluster 5 have the smallest number of tropical cyclones and a lower than all storms landfall percentage. Clusters 2 and 4 both have a smaller landfall percentage than all storms; however, cluster 2 is well populated. Cluster 3 storms do not make landfall in the box chosen.

4.6 Trends

The Poisson distribution is ideally suited to model tropical cyclone counts as they are relatively rare occurring events. The probability of these rare events allowing counts per year (Y_i) with integer values of y=0,1,2, …, mean $\mu(x_i) > 0$, and number of replications for each observation n_i is

$$P(Y_i = y) = \frac{\left(n_i \mu\left(x_i\right)\right)^y}{y!} e^{-n\mu(x_i)} \quad (6)$$

For the Poisson distribution the mean is also equal to the variance. A simple generalized linear model for the mean (variance) that keeps counts positive is a log-linear model

$$\theta = \log \mu(x_i) \quad (7)$$

The global log-likelihood of parameter vector $\theta = (\theta(x_i), \ldots., \theta(x_n))$ is

$$L(\theta) = \sum_{i=1}^{n} \log\left(f\left(Y_i, \theta(x_i)\right)\right) \quad (8)$$

Within a smoothing window, taking advantage of this log link the local log-likelihood function (product of probabilities given by equation 6) is

$$L_x(a) = \sum_{i=1}^{n} w_i(x) \boldsymbol{A}(x_i - x) Y_i \quad (9)$$

where **A** is a vector of explanatory variables (fitting functions), **a** is a vector of the coefficients in a local linear approximation ($a_0 + a_1 x$), x_i are estimated points, x are the observed points, and w_i is the weight

$$w_i(x) = W\left(\frac{x_i - x}{h(x)}\right), W = \left(1 - |u|^3\right)^3, |u| < 1 \quad (10)$$

where h(x) is the bandwidth, and W is the tricube weight function with smoothing parameter u.

The numbers of tropical cyclones per year in each cluster along with the number for the whole basin were analyzed by using this Poisson local linear log-likelihood (Fig. 6.9) by the Matlab version Locfit (Loader, 1999). The fit is made in a moving window (nearest neighbor method) with 70 percent of the data (in this case 41 years) insuring the local neighborhoods always contain a specified number of points (Loader, 1999). The number of years runs along the x-axis, the count on the y-axis with the cluster name as the y label. All clusters show an upward trend in the last two decades, but some are more pronounced than others (clusters 1, 2, 4, and 5). Kossin et al., (2007) found that a trend for an increase in the NS counts in 1984-2004 in the North Atlantic basin was well supported. Goldenberg et al., (2001) also found recent upward trends in intense hurricanes.

The total PDI per year (multiplied by $1*10^{-11}$) is plotted in dots with the Poisson local linear log-likelihood as a line in Fig. 6.10. Year to year the index is highly variable. Emanuel (2007) found the variability to be linked to net surface radiation, thermodynamic efficiency, and average surface wind speed. It appears that the most intense clusters (1 and 5) have the strongest trend in the last two decades, although the last year; 2006; showed a substantial downturn in cluster 1. Elsner et al. 2008 found an increased intensity of the strongest tropical cyclones due to an increase in ocean temperatures over the Atlantic Ocean and elsewhere. Both cluster 1 and the total were raised significantly by Hurricane Ivan in 2004. Ivan not only was a long track, but quickly intensified to categories 4 and 5 and oscillated between the two. The all storms plot differs from Fig. 1 of Emanuel, 2005 (Atlantic basin power dissipation index)

because of the correction Emanuel applied to reduce the wind speed in the pre-satellite era and smoothing (Emanuel, 2005 Supplementary Methods). Landsea (2005) argues that the reduction is unwarranted because "in major hurricanes winds are substantially stronger at the ocean's surface than previously realized".

In summary, the trends for tropical cyclone counts and PDI vary by cluster. The evidence for linear trends over the period of record appears weak. However, there is evidence for an increasing trend in the last decade or so in both variables. The higher activity and intensity in the early part of the record is also notable for several of the clusters. The spatial shift in the PDI trends may be worthy of further investigation.

This page intentionally left blank

Chapter 5

Discussion and conclusions

5.1 Contrasting other North Atlantic clustering studies

This is not the first study to use K-means clustering on North Atlantic tropical cyclones. Elsner (2003) clustered by latitude and longitude coordinates at maximum and final hurricane intensities and found three clusters. Elsner's study did not include TS tracks or portions of tracks resulting in shorter trajectories. Figure 1 of that paper shows the three clusters. Elsner finds the shorter trajectory straight moving cluster (red) to be the most intense. By including the tropical storm part of track, we found the longest tracks to be the most intense (correlation of 0.68 for length and intensity of all clusters; which is a significant t value for a two-tailed distribution at the 0.01 level). The three clusters are similar to the results we found when only clustering centroids (Fig. 6.11 top); the cluster regimes have nearly N-S oriented breaks.

Camargo et al., (2008) has also completed preliminary clustering in the North Atlantic incorporating the whole track length and actual location by using a regression mixture model on the entire HURDAT dataset. Three clusters were found: Gulf, Cape Verde, and Eastern US born storms (11 bottom). When the K-means analysis is re-done on the

similar time period (1851-2003) for centroid and variance, the resulting clusters are strikingly similar (Fig. 6.11 middle). However, in this method, silhouette values indicate the optimal number of clusters to be six. In their North Pacific paper (Camargo et al., 2007) they found a similar result of longer cyclone life spans leading to more intensification. Cluster number in these papers is based on estimation of the diminishing return of log-likelihood plots and visual analysis. Camargo et al. (2008) also note that their finite mixture model produces better results than K-means because they take into account the whole track shape and location. Moments allow the use the simple and computationally elegant K-means while also taking into account the whole track shape and location.

5.2 Summary

A novel approach for clustering storm tracks was developed using K-means clustering with the mass moments of centroid and variance. The centroid captures where the tropical cyclone is located and the variance describes the entire track shape. A summary of results are as follows:

- A tropical storm track can be thought of as an open curve allowing computation of the first and second moments (centroid and variance) (Fig. 6.1).
- K-means clustering applied to the first two moments for each track provides a method for selection of clusters or groups. The group number is selected through the mean and number of negative silhouette values (Fig. 6.2).

- The resulting clusters (Fig. 6.3 and Table 6.1) have not only grouped centroid locations, but also distinctly differently shaped variance ellipses.
- Although genesis locations of some clusters are diffuse (clusters 2, 3, and 5 – see Table 6.2) the track shapes (Fig. 6.4) clearly follow a pattern: cluster 1 is the classic parabola, 2 are Eastern US Coast storms, 3 are the flattened parabolas, 4 are the straight moving, 5 are the semi-parabolas, and 6 are the Gulf storms.
- Both maximum wind speed and intensity were found to be linked to life span (correlations of 0.59 and 0.68 respectively over all clusters; which is significant for a two-tailed distribution at the 0.01 level) with longer tracks producing higher maximum wind speeds and more intense tropical cyclones (Figs. 6.5, 6.6, and 6.7).
- Half of the clusters showed the typical four month long Atlantic tropical storm season of July to October, while two clusters had a longer six month season of June to November, and one cluster showed an eight month season of April to November (Fig. 6.8).
- Landfall percentages of the clusters were clearly different ranging from 0% to 85% (Table 6.4).
- Trends for the number of cyclones per year show an upward trend in the last two decades, but some are more pronounced than others (Fig. 6.9).
- Upward trends maximum intensity of the most intense cyclone clusters was found to be stronger than those for all storms (Fig. 6.10).

- Results of Elsner (2003) K-means study using two points along the track appear to correspond well to those obtained by only using the first moment of the tracks in K-means (Fig. 6.11 top). Thus, Elsner (2003)'s method uses only part of the information.
- Results of Camargo et al. (2008) regression mixture model are similar to those obtained using both the first and second moments in K-means (compare Fig. 6.11 middle and bottom).

Future work will include exploring the meteorological conditions behind the six clusters. As tropical cyclone formation and movement is based on meteorological conditions the hypotheses of clusters also being related is a sensible one. If clear connections are found then known meteorological conditions can be linked to a particular class and shape of storms. Possible risk assessment applications include relating climate fields or indices to an increased or decreased chance of tropical cyclone genesis, genesis location, and track shape. These in turn can be related to landfall and possible landfall intensities, which is of interest to government planning boards and insurance companies.

This page intentionally left blank

Chapter 6

Figures and Tables

Figure 6.1: Example of centroid (star) and variance ellipse for track. Track start marked with circle.

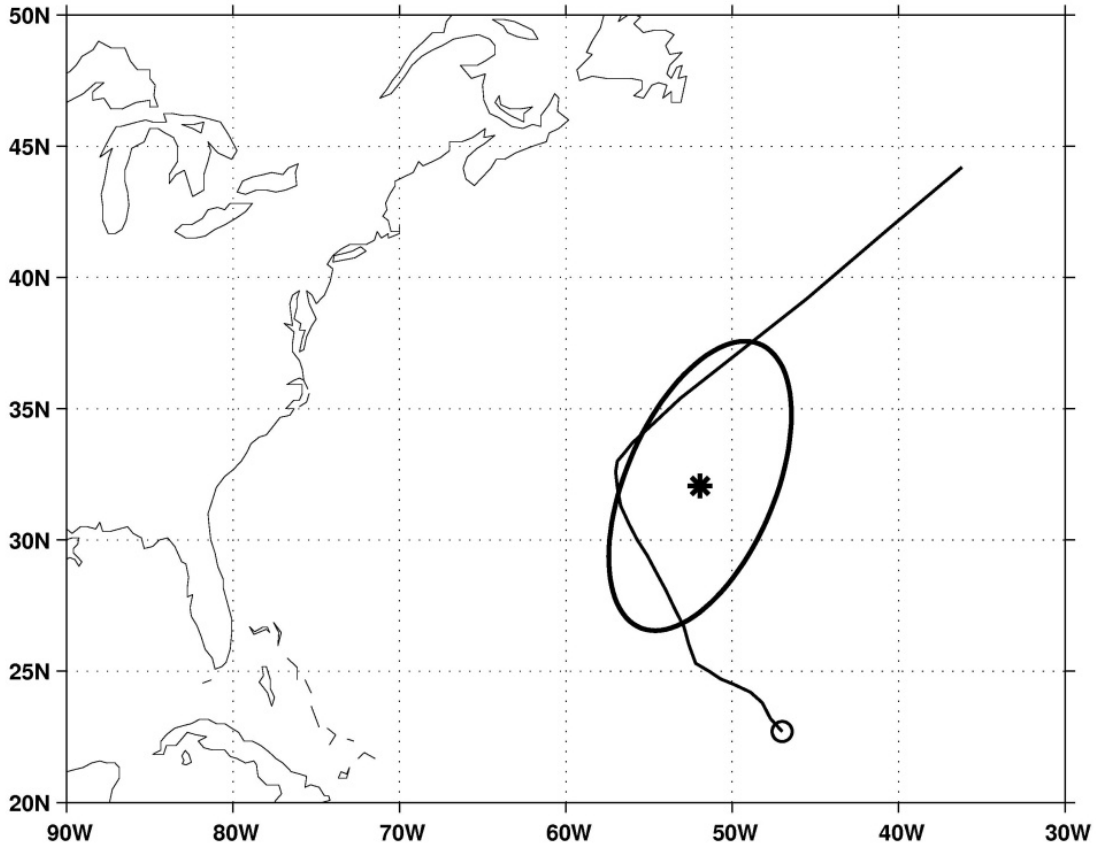


Figure 6.2: Mean silhouette values (top) and number of negative silhouette values (bottom).

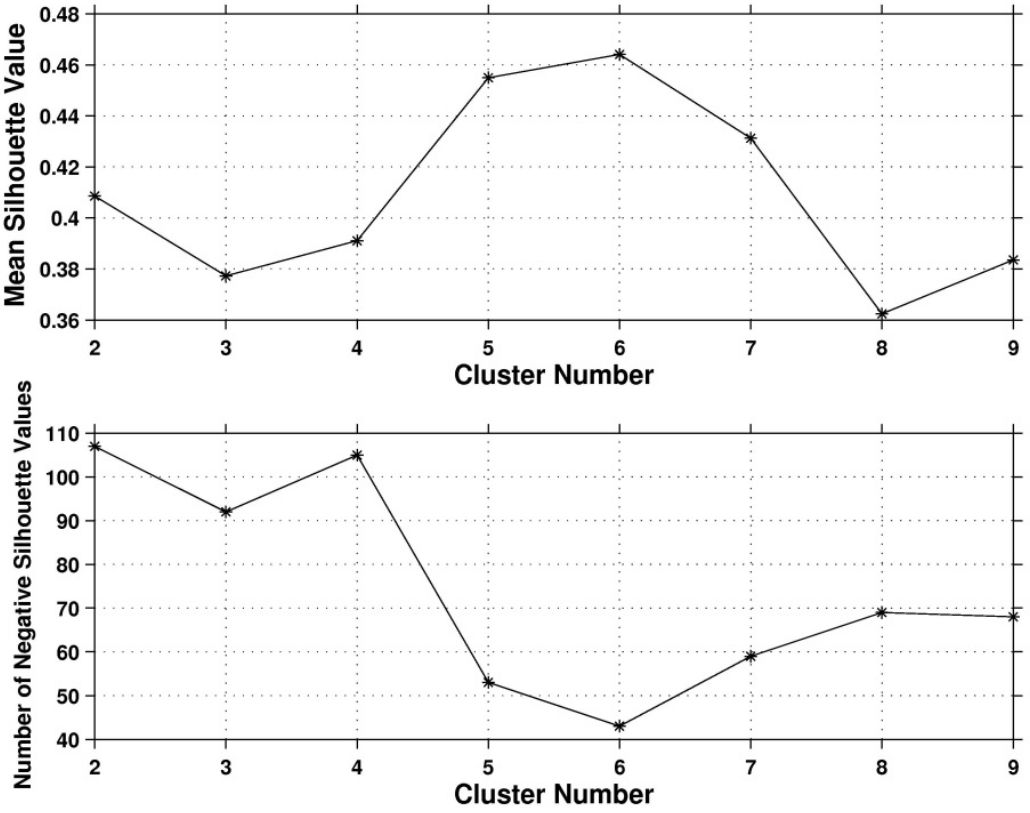


Figure 6.3: Centroid locations (stars) and directional variance (light ellipses) for the six clusters. Mean centroid value is marked with a dark x, and mean variance ellipse with a dark line.

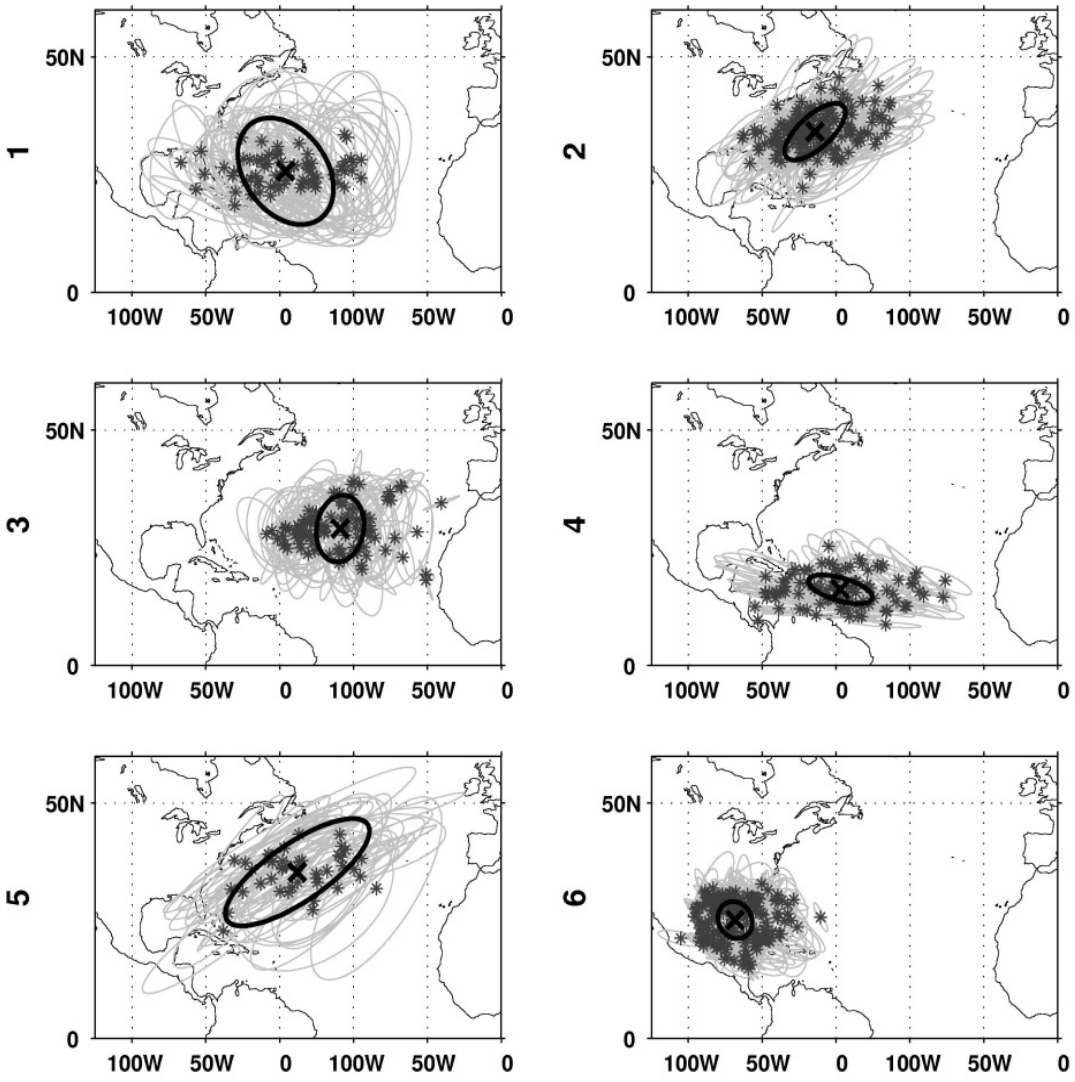


Figure 6.4: Genesis location (circle) and the track (line) for the six clusters.

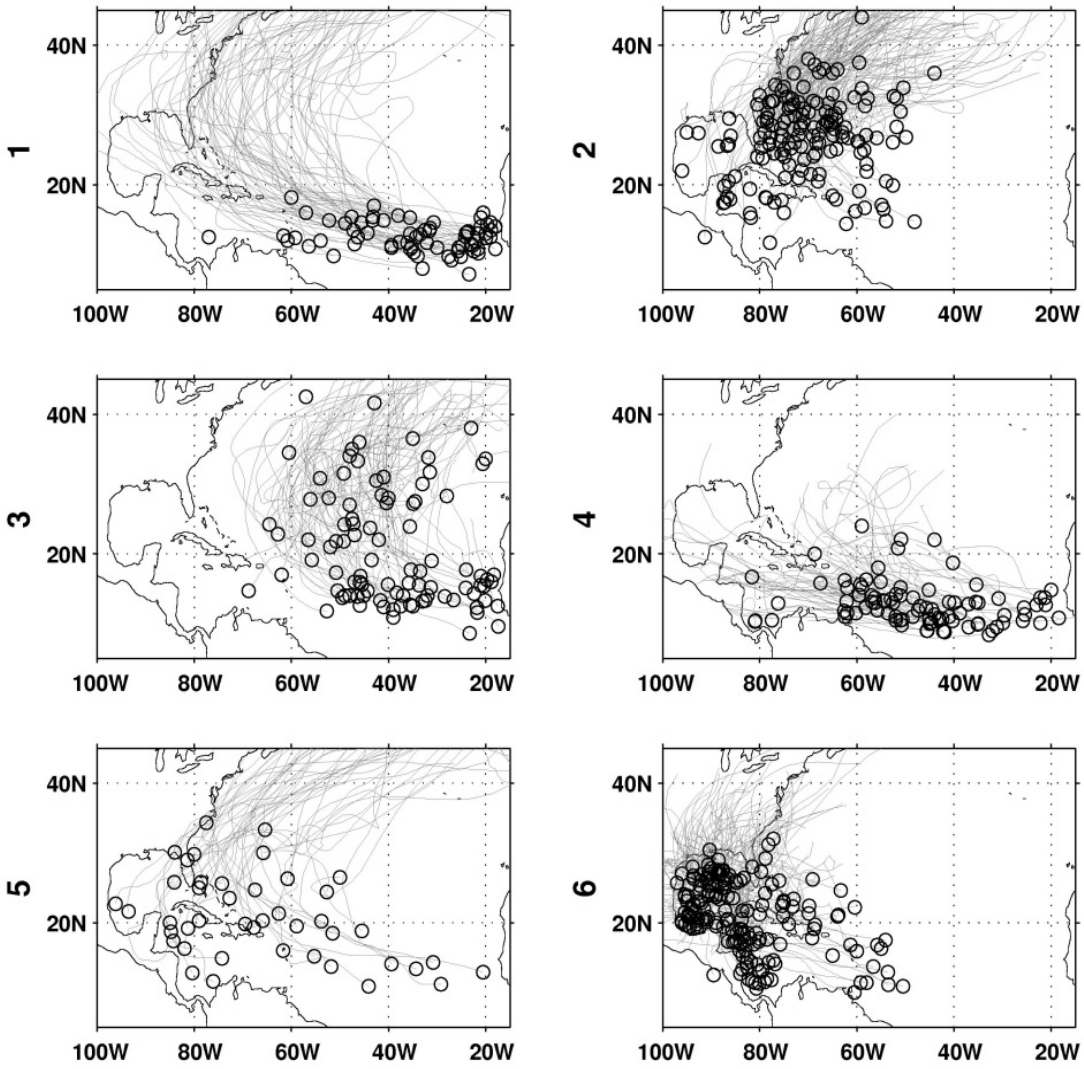


Figure 6.5: The 25th and 75th percentile (upper and lower bounds of the box), the mean (star), the median (bar in middle), the bounds (dashed line), and outliers (plus signs) of the distribution of tropical cyclone maximum wind speed in each cluster and as a whole.

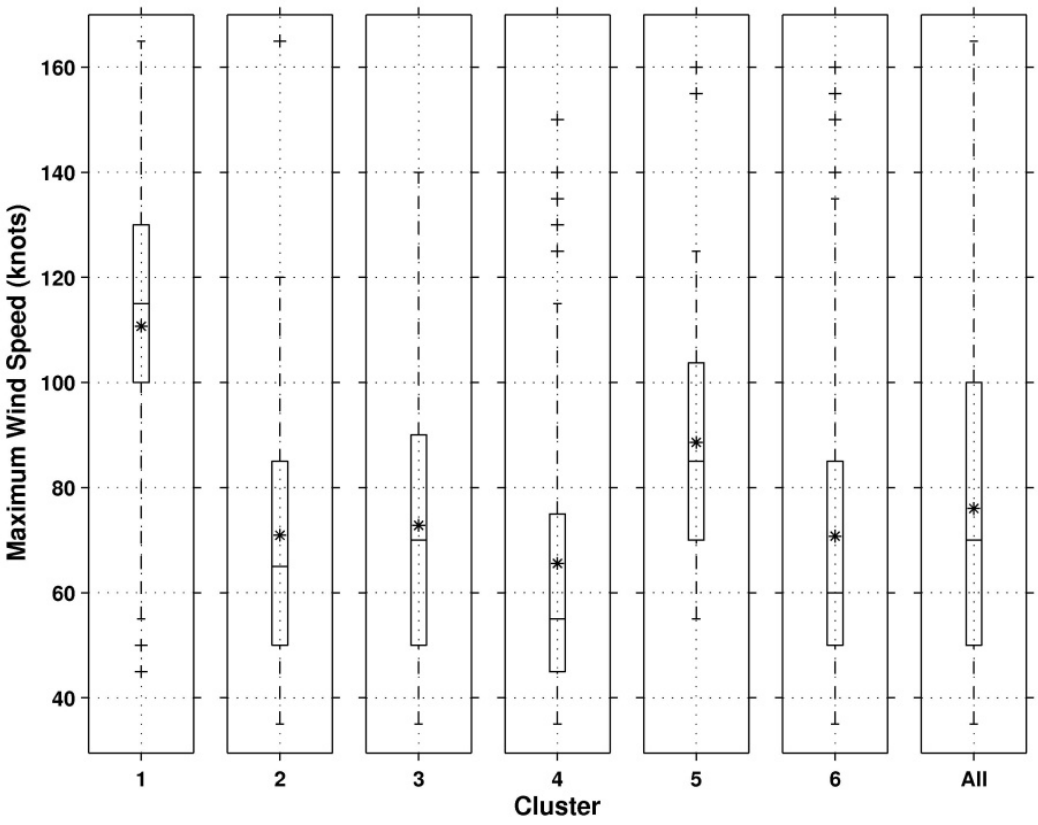


Figure 6.6: The 25th and 75th percentile (upper and lower bounds of the box), the mean (star), the median (bar in middle), the bounds (dashed line), and outliers (plus signs) of the distribution of tropical cyclone lifespan (days) in each cluster and as a whole.

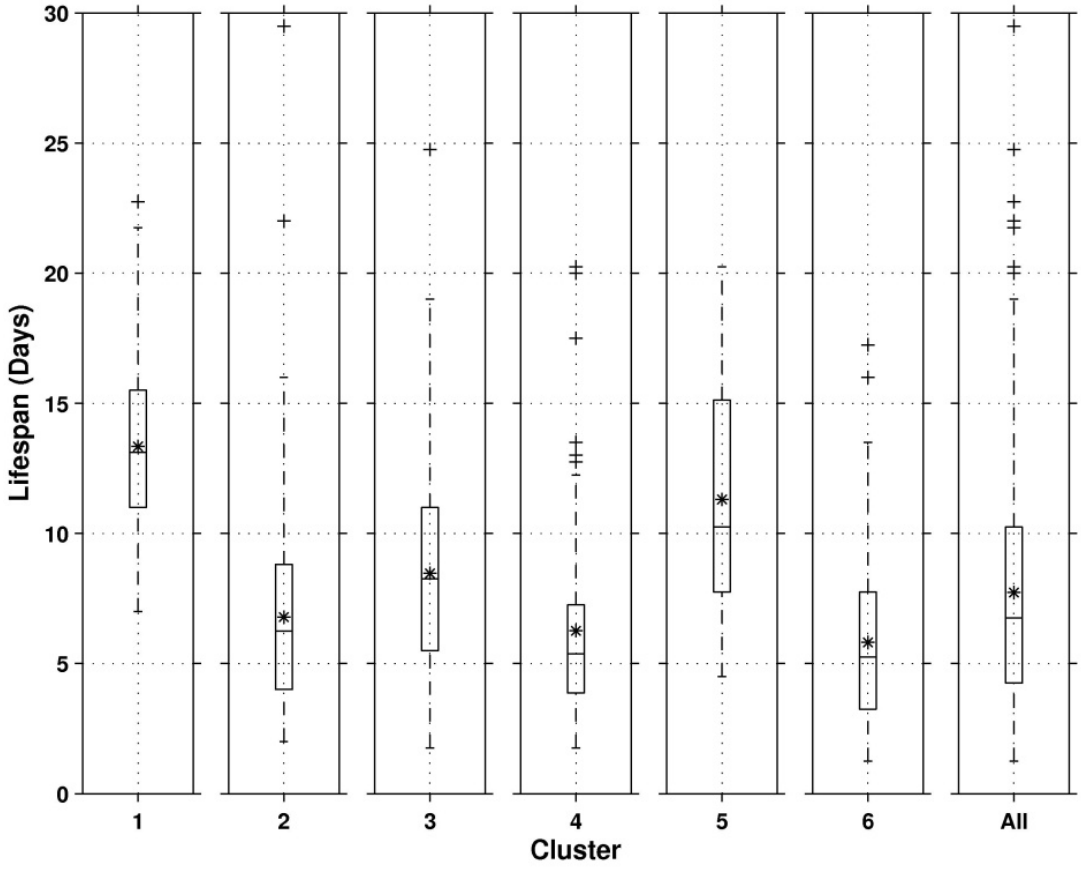


Figure 6.7: The 25th and 75th percentile (upper and lower bounds of the box), the mean (star), the median (bar in middle), the bounds (dashed line), and outliers (plus signs) of the distribution of tropical cyclone PDI multiplied by $1*10^{-11}$ in each cluster and as a whole.

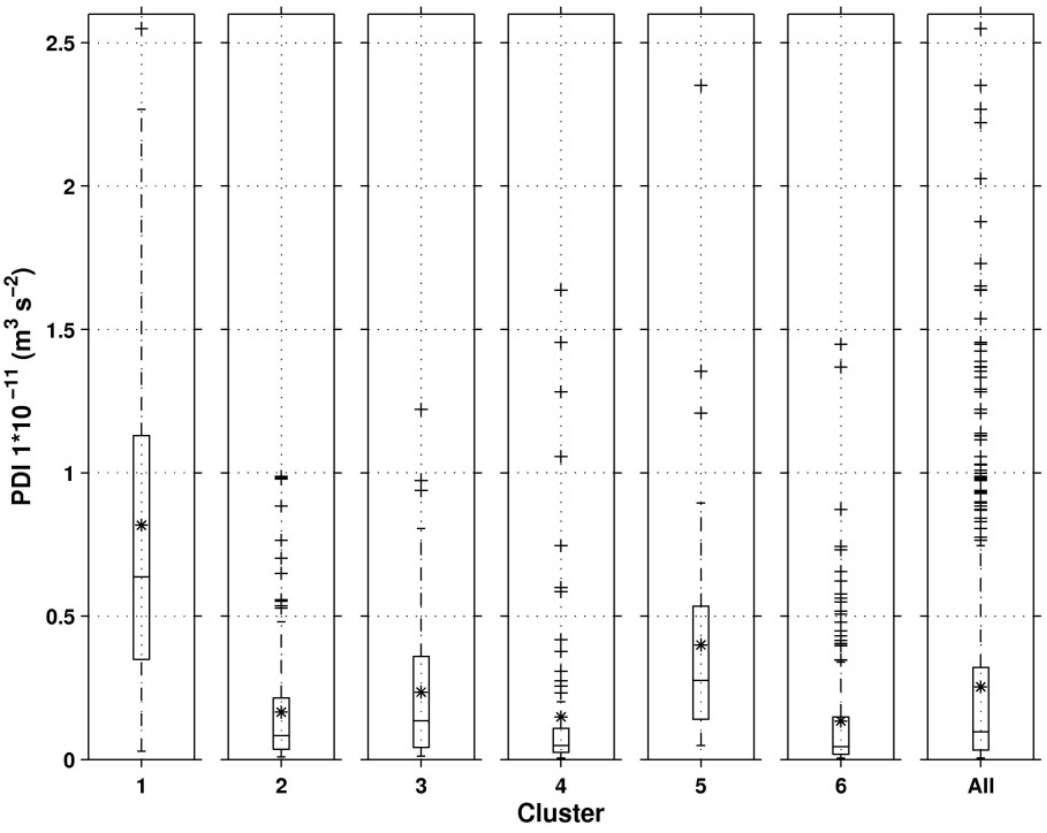


Figure 6.8: The 25th and 75th percentile (upper and lower bounds of the box), the mean (star), the median (bar with circle in middle), the bounds (dashed line), and outliers (plus signs) of the distribution of tropical cyclone month in each cluster and as a whole.

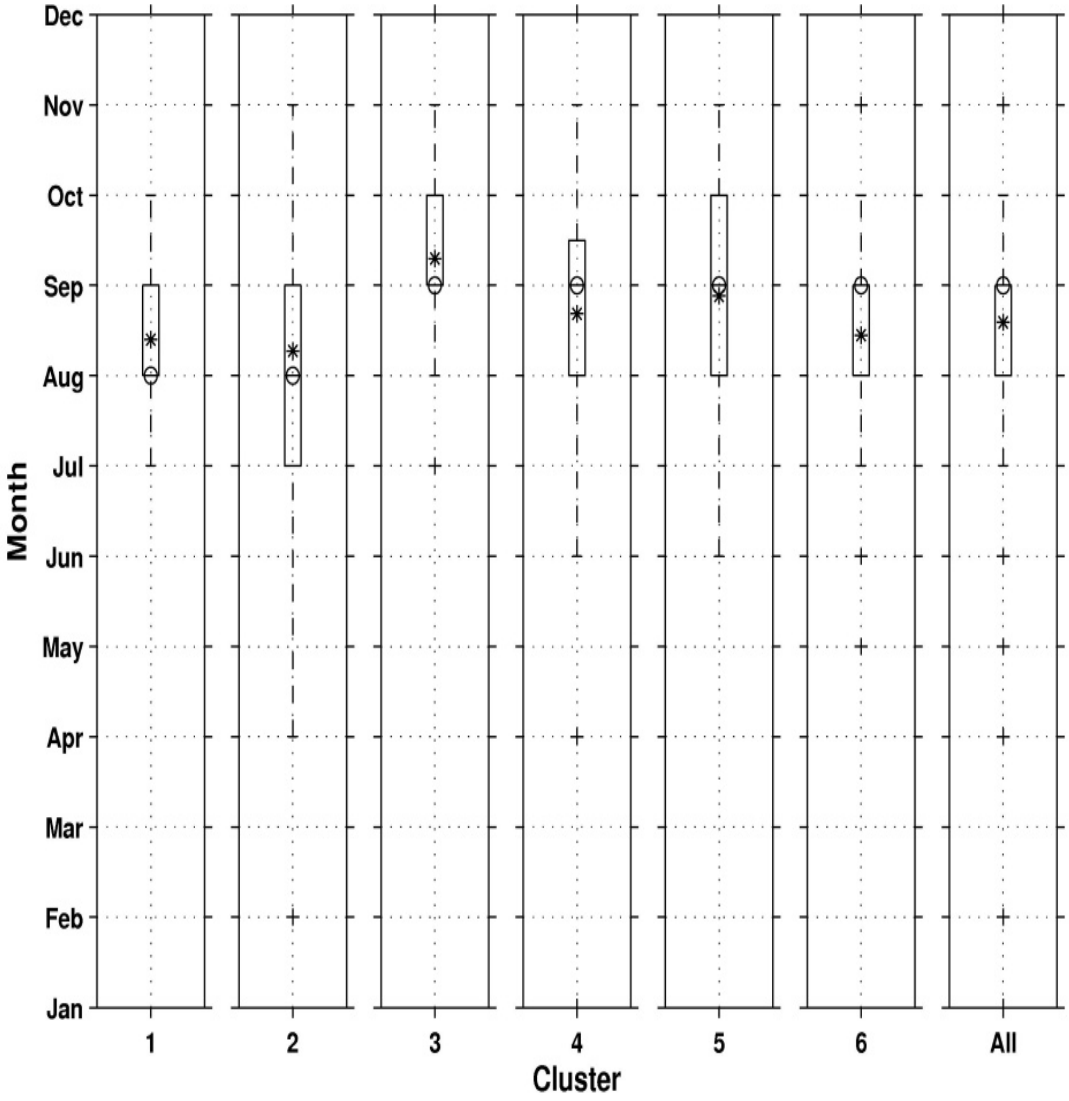


Figure 6.9: The number of tropical cyclones per year in each cluster along with the number for the whole basin (dots) with the linear Poisson local linear log-likelihood fit as a line.

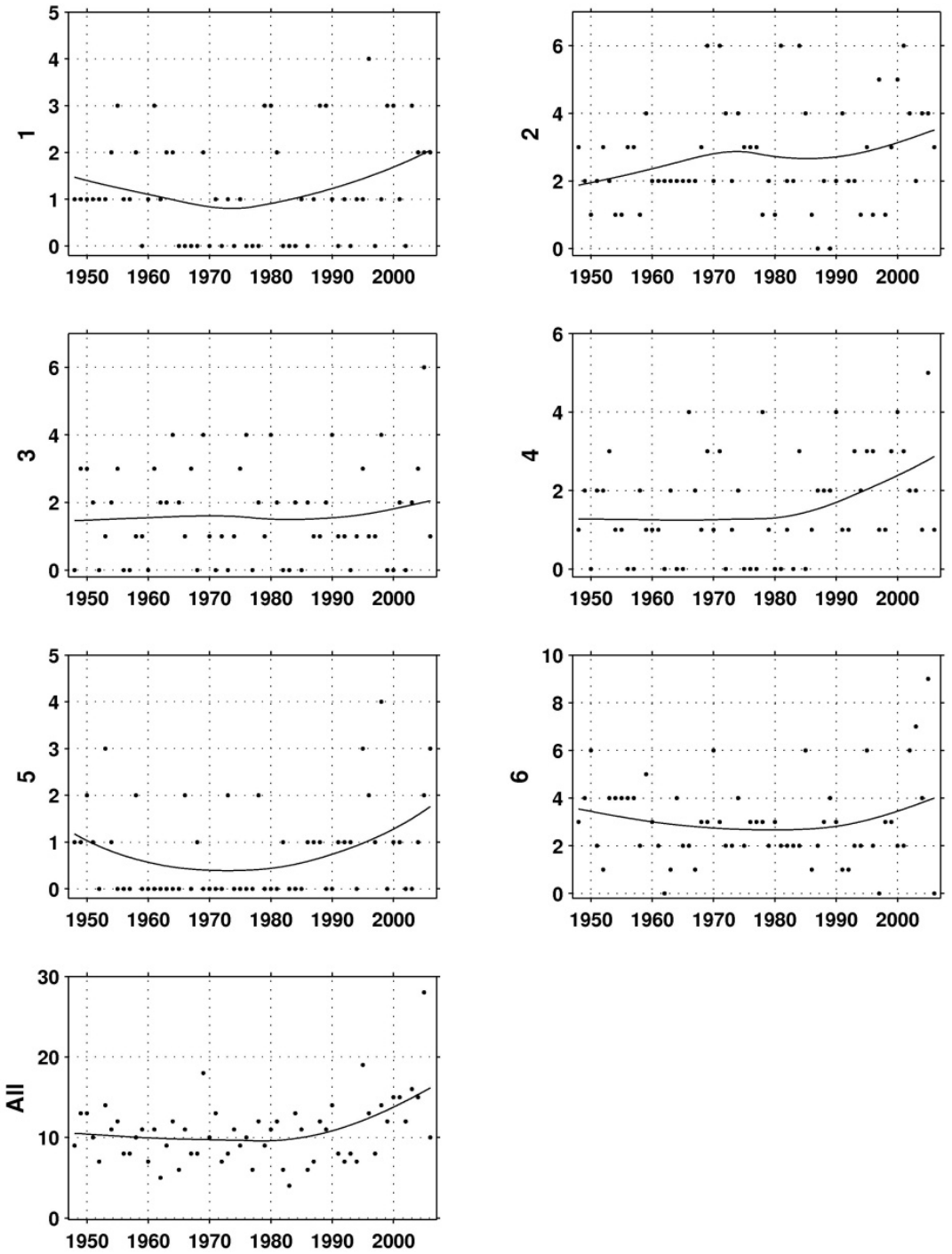


Figure 6.10: PDI (m^3 s^{-2}) per year multiplied by $1*10^{-11}$ in each cluster along with the number for the whole basin (dots) with the linear Poisson local linear log-likelihood regression fit as a line.

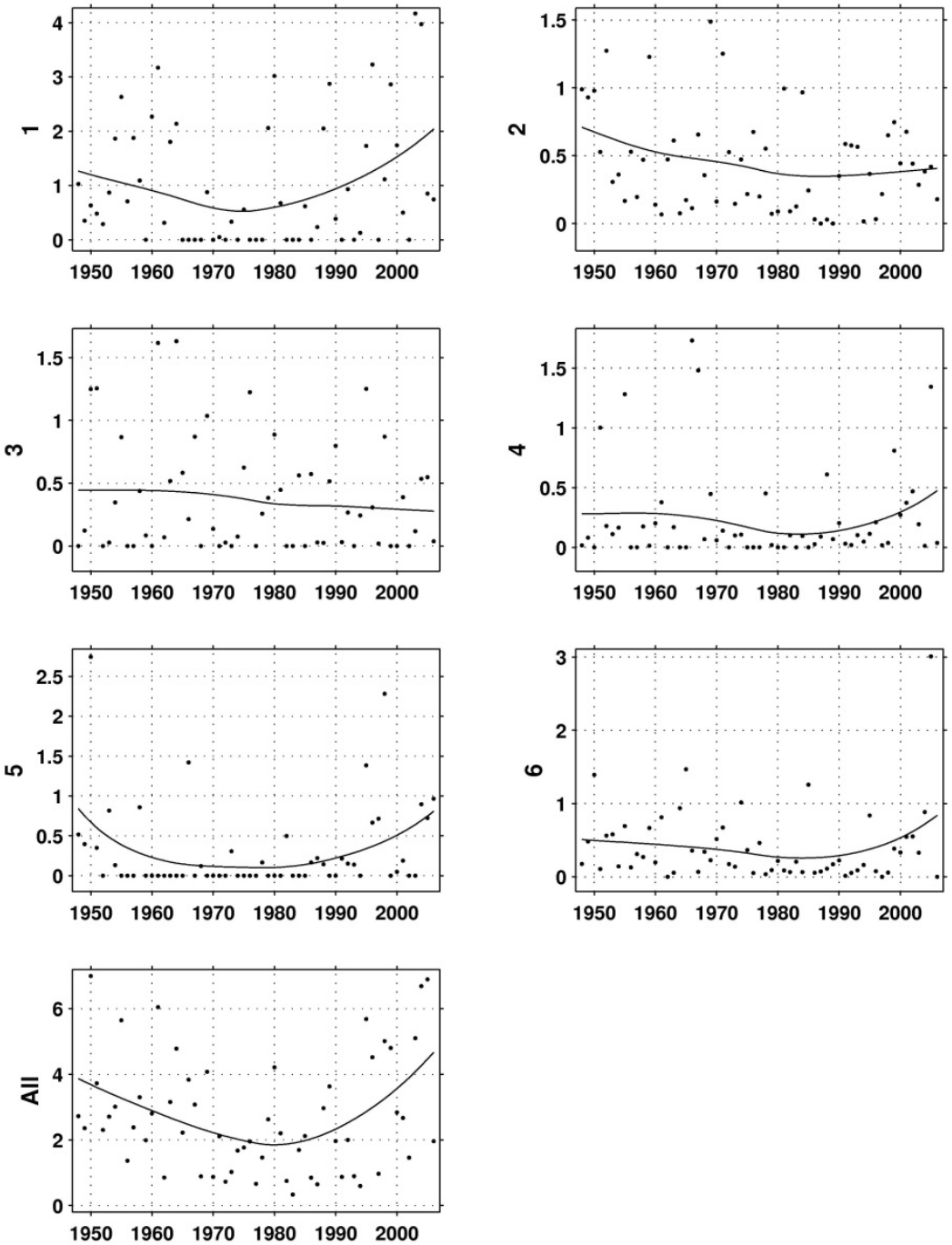


Figure 11: Centroid locations of three cluster 1851-2003 (Top) K-means on centroids only, (Middle) K-means on centroid and variance, and (Bottom) Camargo et al., (2008) regression mixture model.

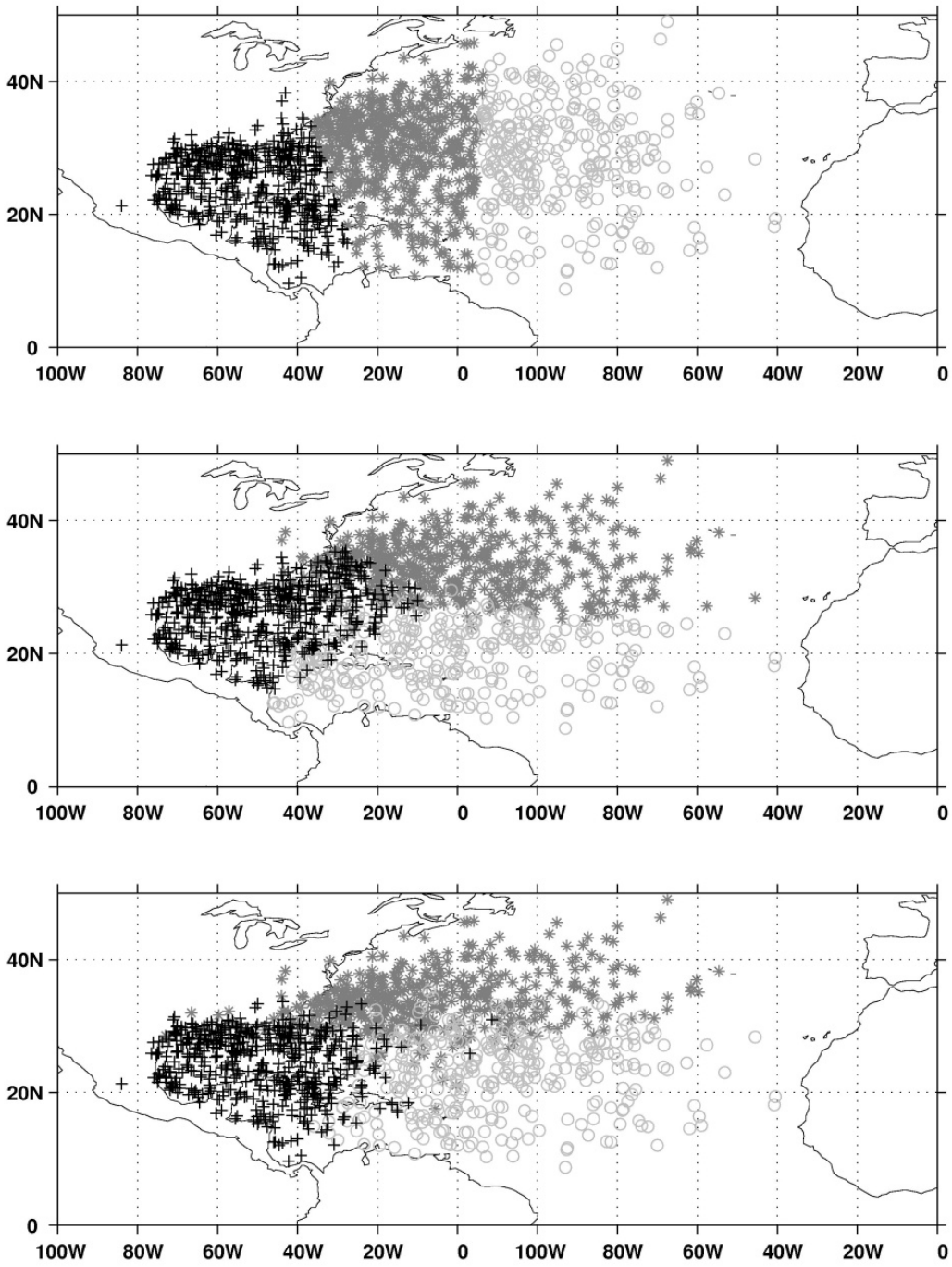


Table 6.1: Mean centroid and variance values for each cluster

Cluster #	Centroid X	Centroid Y	Variance X	Variance Y	Variance XY
1	58.38W	25.69N	168.25	128.29	-47.89
2	65.68W	34.07N	67.20	36.33	33.99
3	43.06W	29.06N	41.89	50.48	5.64
4	58.83W	16.16N	75.28	9.34	-14.98
5	55.20W	55.20N	381.27	129.95	171.38
6	87.32W	25.15N	21.74	15.57	-2.48

Table 6.2: Mean and range of tropical cyclone genesis locations for each cluster

Cluster #	Mean X genesis	Mean Y genesis	Genesis X range	Genesis Y range
1	34.60W	12.61N	77W to 18W	7.2N to 18.2N
2	70.50W	26.64N	96W to 44W	11.7N to 44N
3	39.35W	20.76N	68.8W to 17.4W	8.6N to 42.5N
4	46.83W	12.76N	81.6W to 14W	8.4N to 24N
5	65.89W	20.67N	96.2W to 20.5W	10.9N to 34.3N
6	83.16W	20.77N	97W to 50.5W	10N to 32N

Table 6.3: Mean and median category of the maximum wind speed in each cluster and in all tropical cyclones.

Cluster #	Mean Category	Median Category
1	3	4
2	1	1
3	1	1
4	1	TS
5	2	2
6	1	TS
All	1	1

Table 6.4: Number of tropical cyclones, landfalls, and landfall percentage for each cluster and for all tropical cyclones

Cluster #	Number of Cyclones	Landfalls	Landfall Percentage
1	70	32	46%
2	157	35	22%
3	94	0	0%
4	92	26	28%
5	43	14	33%
6	174	148	85%
All	630	255	40%

This page intentionally left blank

Part II

Hurricane Intensity and Track Simulator (HITS)

with Atlantic Ocean Applications

for Risk Assessment.

This page intentionally left blank

A nonparametric stochastic model is developed and tested for the simulation of tropical cyclone tracks. Tropical cyclone tracks demonstrate continuity and memory over many time and space steps. Clusters of tracks can be coherent, and the separation between clusters may be marked by geographical locations where groups of tracks diverge due to the physics of the underlying process. Consequently, their evolution may be non-Markovian. Markovian simulation models, as often used, may produce tracks that potentially diverge or lose memory quicker than nature. This is addressed here through a model that simulates tracks by randomly sampling track segments of varying length, selected from historical tracks. For performance evaluation, a spatial grid is imposed on the domain of interest. For each grid box, long-term tropical cyclone risk is assessed through the annual probability distributions of the number of storm hours, landfalls, winds, and other statistics. Total storm length is determined at birth by local distribution, and movement to other tropical cyclone segments by distance to neighbor tracks, comparative vector, and age of track. An assessment of the performance for tropical cyclone track simulation and potential directions for the improvement and use of such model are discussed.

Citation: Nakamura, J., U. Lall, Y. Kushnir and B. Rajagopalan, 2014: HITS: Hurricane Intensity and Track Simulator with N. Atlantic Ocean applications for risk assessment. *Submitted J. Appl. Meteorol. Climatol.*

This page intentionally left blank

Chapter 7

Introduction to Part II

Extreme weather events such as tropical cyclones occur with low frequency. Due to the low probability of North Atlantic tropical cyclone landfall, landfall statistics are difficult to estimate. Statistical methods can be employed to resample the historical data, creating a large numbers of tracks used to improve estimates of the probability extremes. These statistics are useful for insurance companies in determining premiums, and for communities and government in determining disaster plans and building codes.

Several models have been introduced to estimate tropical cyclone landfall probabilities. Some directly simulate tropical cyclone landfall and intensity by Monte Carlo methods (Clark 1986, Chu and Wang 1998), dimensionality reduction (Buckman 2011), or by fitting a probability function (Emanuel and Jagger 2010), while others create sets of simulated tracks (Casson and Coles 2000, Emanuel et al. 2006, Hall and Jewson 2007 and 2008, Hallegate 2007, Rumpf et al. 2007 and 2009, Vickery 2000, Yonekura and Hall 2011). Casson and Coles (2000) simulate tracks by the shift of a uniform random value less than 100 nautical miles and use a simulated pressure plus land effects to produce a simulated wind speed. Emanuel et al. (2006) present a smoothed Markov

chain model and beta and advection model using synthetic flow at 850 and 250 mb. Hallegate (2007) used the later to investigate landfall and damage probabilities. Hall and Jewson (2007 and 2008) and Yonekura and Hall (2011), model latitude and longitude displacements separately based on the mean and variance of the local area plus random error. Rumpf et al. (2007 and 2009) group tracks and use histograms of local area direction, speed, and wind speed of tracks to create simulated tracks. Vickery (2000) used a 5-degree by 5-degree box to determine track heading, speed, and intensity, through a regression of historical storm data on location attributes, prior time step storm speed and direction. Hurricane intensity is modeled as a function of prior intensity for up to 3 time steps, and of sea surface temperature at appropriate locations. A random innovation consistent with the regression model is added to generate the tracks. A number of other parameters are also estimated in a regression framework, and landfall probability distributions are estimated from the simulations. The variance explained by their regression models is typically not high, but the conditional probability distributions derived generally look plausible. In summary, the simulation models presented in the literature are by and large Markovian, looking at ways to march the track one step (in time and/or gridded space) at a time based on conditional probabilities inferred from track and climate attributes. The more successful models appear to consider transition probabilities or conditional probability distributions that change with spatial location and with respect to exogenous attributes, but not with respect to the age of the track, or the extended past history of the track.

To a first order approximation, tropical cyclones move in the direction that the winds (over the depth of the storm) steer them. In the Northern Atlantic, the northeasterly trade

winds move the storms westward from the African coast. The prevailing flow around the subtropical high curves them, and other cyclones generated in the Caribbean and Gulf of Mexico, northward approaching the North American coast and then eastward in the middle latitudes. Elsner and Kara, (1999) call this a parabolic sweep. The position and strength of the subtropical high, the extra-tropical circulation, and the birth location of the tropical cyclone varies, allowing variation in tracks. To reflect the parabolic sweep, and variations of it, a spatial model is needed. More "popular" track paths have a higher probability of occurring than "unpopular" ones.

In our work in the late 1990s, a Markov chain model was initially implemented to simulate these popular paths with transition probabilities changing from grid box to grid box, but it was found to be too diffusive. For each grid box, the one step transition probability to adjoining grid boxes and to the grid box (5 degree) itself was estimated. The Markov chain model was fit to the post airplane reconnaissance historical record (1944-1999) and then track simulations were generated seeded by a random selection of historical track birth positions. A comparison of the historical storm residence time in each grid box with that simulated is presented in Fig. 12.1. The match is rather poor, reflecting the diffusion one qualitatively expects from a "lag one" space-time model, relative to a model that considers "runs" or longer-term persistence. Considering higher dependence (e.g., to the 2 prior steps) in a Markov chain model, leads to an explosion in the number of parameters to be estimated for the resulting transition probability matrix, which is commonly called the "curse of dimensionality" and is hence not indicated with the hurricane track data set. In the time series modeling literature, the situation is often addressed by considering a semi-Markov or Markov Renewal model (Bhat and Miller,

1972; Cinlar, 1969, 1975; Gilbert et al, 1972, Foufoula-Georgiou and Lettenmaier, 1987). A direct application of the Markov Renewal model to the hurricane track setting is not obvious at first glance, since one needs a specification of discrete states for the system, prior to modeling the conditional distribution of the time to be spent in each future state. However, inspired by the Markov Renewal idea, we propose a modeling strategy, where we consider the time t_i to be spent along a candidate track i to be a random variable, and allow the selection of the candidate track and the associated t_i to depend on location and other attributes. This provides the basis for the hurricane interactive track simulator (HITS) model presented in this paper.

The historical hurricane record is discussed in chapter 8, along with 2012 data used for model verification on novel tracks. Chapter 9 describes the HITS algorithm and chapter 10 presents the results of the comparison to the historical record using percentiles, and a visual comparison of distributions with split violin plots. Chapter 11 discusses results, a brief comparison to other hurricane track models, and plans for future work.

This page intentionally left blank

Chapter 8

Data

The data used for the model are the historical best track North Atlantic hurricane data (Jarvinen et al. 1984) from 1851 to 2011 with information on storm position (latitude/longitude) and wind speed every six-hours (HURDAT). The data were obtained from the National Hurricane Center, and the record contains "named" storms, defined as storms whose maximum sustained (one min averaging period) surface winds exceeded 17 m s^{-1} (34 knots, 39 mph). The primary table of tropical cyclone data consists of storm number that year, catalog storm number, year, month, day, position on the track (six-hour time step), latitude, longitude, and wind speed (knots). A domain of 5N to 45N and 100W to 25W was chosen and segmented by a 5° by 5° grid to give 120 boxes. These grids are used to report model performance evaluation. The model is built directly using the track data without spatial discretization. During the 161-year record, 1465 tropical cyclones spent all or part of their lifetime in the domain.

Due to routine aircraft reconnaissance missions into tropical cyclones beginning in 1944, details on the position of the tropical cyclone eye have been available. This has led to greater accuracy in the six-hour position data in storms far from land or shipping lanes. The distributions of storm duration in the period prior (1851 to 1943), and subsequent to

(1944 to 2011) this change in observing method are not equal in length, with shorter tracks in the prior period. Birth location of the storms is sampled from the post 1944 data by necessity. However, all available track data is used for neighbors to model tropical cyclone movement behavior, following a philosophy of even non-ideal data inclusion (Halevy et al. 2009).

This page intentionally left blank

Chapter 9

HITS Algorithm

The basic modeling premise in HITS for hurricane track simulation is based on the following conceptual model. Consider a location, **x***, and an area A($\mathbf{x}^*$) centered at **x***. Tropical cyclone tracks that pass through A($\mathbf{x}^*$) describe random curves in space, whose evolution is conditional on specific attributes of the location **x*** and of the track history. Consider a track or curve C(**x***) that emanates from **x***. Its evolution in space and time may then be represented through a conditional probability distribution:

$$f(C(\mathbf{x}^*), t(C(\mathbf{x}^*)) | q(\mathbf{x}^*))$$

where by analogy to a Markov Renewal model, we may index the state variable, C(**x***) to correspond to one of the historical tracks, j, and use t(C(**x***)) to represent the random time to be spent in that state starting from the location **x***. Thus, given a certain location, the model considers the selection of a state (track) going forward from that location, and the time that is to be spent following that track. This permits one to consider persistence of motion along tracks and avoids the diffusion associated with the one-step Markov chain models.

The q(**x***) is then a set of parameters relevant to the tropical cyclone process that vary by location. The q(**x***) may include for instance, the location coordinates **x***, the index and vector of the track C(**x***)=i, that was traveled to reach the location **x***, atmospheric variables that influence track selection, and large scale climate variables, such as sea surface temperatures or indices such as the NINO3.4 or NAO. The specific choices for q(**x***) made in the current application, and the nonparametric estimation of the conditional probability density function f(C(**x***), t(C(**x***))|q(**x***)) used to simulate the tracks are discussed below as part of the algorithm presentation.

The HITS algorithm for simulating tropical cyclone tracks steps through time in six-hour intervals, making decisions along the path as to which historical track to follow. This process repeats until the lifetime is met. A schematic of the process is provided in Fig 12.2.

Each simulated hurricane season the following steps are taken:

(1) Selection of number of storms in a season.

$$N \sim U[N_l, N_u]$$

where N is the number of storms in a simulated season, N_l is the lowest number, and N_u is the highest number of storms in years 1944 to 2011, and U[] is a uniform draw. This can be conditioned on large-scale climate state to reflect the dependence of number of tropical cyclone births on ENSO or other climate state. However, in the work presented here, we do not consider such a dependence.

(2) For each potential storm, randomly select a birth location x_b from all historical births or historical births for years corresponding to a specific condition (El Nino, position of the Bermuda high, etc.)

$$x_b \sim U[S(x_b)]$$

where $S(x_b)$ is a set of all candidate birth locations (latitude and longitude), and x_b is the simulated track birth location. Again, in the applications presented in this paper we choose the birth location unconditionally from the candidate locations.

(3) Sample simulated track lifetime:

$$L \sim U[L_1,L_2]$$

where L is the entire duration of the cyclone, and L_1 and L_2 are the minimum and maximum total lifetimes of the tracks that lie in $B(x_b)$ where $B(x_b)$ is an area with a 2.5° radius around x_b, Tracks recorded in the years 1944 to 2011 are considered.

(4) Define new position on track i

$$x_j = x_b + S_{ij}$$

where x_j is the position on the simulated track at iteration j, x_b is the birth location, and S_{ij} is the number of time steps represented by t(C(x*)), the random amount of time in this state, to take along chosen track i moving forward from x_j. t(C(x*)) ~ U[t*,L], where t* is the time elapsed on the current track, and L is the simulated track length.

(5) Clustering results of N. Atlantic tropical cyclone tracks have shown six groupings that display differing genesis locations, track shapes, intensities, life spans, landfalls, seasonal patterns, and trends (Nakamura et al. 2009). The selection of the probability criteria (q(**x***)) of distance, vector difference, and age (step number) were selected to statistically capture and display these dynamical behavior differences of tropical cyclones born in different parts of the basin through the conditional probability density function f(C(x*), t(C(x*))|q(x*)).

Choose a track in years 1851 to 2011 by drawing form "neighbor" tracks using the conditional density function defined through a product kernel density function as

$$f(C(x^*),\tau(C(x^*)\,|\,\theta(x^*)) \propto k(u_{1ij})k(u_{2ij})k(u_{3ij})$$

where

$$k(u) = (1-u^2)^2$$

is the bisquare kernel function, and u_1, u_2, and u_3 represent distance measures in terms of different conditioning variables, as described below

a) Distance of a historical track to the current track

$$u_1 \propto \frac{D}{2.5 * \frac{\pi}{180}} \qquad D \le 2.5^{\circ}$$

where D is the distance between neighbor track points in spherical geometry.

b) Orientation of a historical track relative to the current track

$$u_2 \propto \frac{V}{Max(V)} \qquad V = \left(V_{xn} - V_{xi}\right)^2 + \left(V_{yn} - V_{yi}\right)^2$$

where Max(V) is the maximum wind vector difference over all tracks, V_x is the instantaneous storm maximum wind speed in knots times the cosine of the angle between the current point and the next point on the track, V_y is the wind speed in knots times the sine of the angle, subscript i is the current track and subscript n is the neighbor track.

c) Age of the historical track relative to the current track:

$$u_3 \propto \frac{T}{Max(T)} \qquad T = T_n - T_i$$

where T is age of the neighbor track minus the age of the current track, Max(T) is the maximum T over all tracks, subscript i is the current track and subscript n is the neighbor track.

The final probability moving to neighbor f(C(**x***), t(C(**x***))|q(**x***)) is proportional to the product of the distance, comparative vector, and track age probabilities.

(6) The steps remaining to the simulated storm end (R_j)

$$R_j = L - a_j$$

where L is the duration selected in step 3, and a_j is the age (number of steps taken) at iteration j.

(7) Steps to take along chosen track i (S_j)

$$S_j \sim U[1,R_j]$$

where R_j is selected in step 6.

(8) Position on simulated track at iteration j

$$x_j = x_j + S_j$$

(9) If $S_j = R_j$ then stop, else repeat steps 5 to 8

(10) Repeat steps 1 to 9, N (selected in step 1) times to complete a simulated season.

A numerical procedure for the efficient sampling and simulation of tracks was developed. A functional table was created that recorded the storm number, position on the track, latitude, longitude, number of time steps to the end of the track, the comparative vector of storm direction and wind speed. A second lookup table was created that listed the "neighbors" for all tracks for each position in each track. Neighbors are all points on other tracks within a 2.5° radius of the current location with a given probability based on the bisquare kernel function to move to that location based on distance, comparative vector, and age in six-hour steps as defined in step 5.

As an example, Fig 12.3 shows the actual 2011 North Atlantic hurricane season tracks (top), and a HITS algorithm simulation of an arbitrary hurricane season plotted at the bottom. Fig 12.3b shows that the tracks have more variation than the historical tracks, but they are still more coherent than the tracks from a Markov chain model. Fig 12.3b also illustrates the jump discontinues of the simulated tracks as they move to neighbors; however, the accuracy of the model is assessed on the binned statistics in chapter 10 rather than the track paths.

This page intentionally left blank

Chapter 10

Results

The HITS algorithm described in the previous section was applied to the Atlantic basin tropical cyclone data, and 1000 hurricane seasons were simulated. The performance of the simulations is judged through a variety of performance measures:

10.1 Comparing the average spatial distribution of the historical and simulated data

10.2 Comparing percentiles of the residence time in each grid box

10.3 Landfall statistics

10.4 Comparing frequency of six-hour periods of hurricane strength wind (>64 knots)

In addition to these measures, we explore the conditional probability density of hurricane tracks simulated from selected starting positions for the 2012 storms (Sandy and Isaac) whose data were not included in model building.

10.1 Comparing the average spatial distribution of the historical and simulated data

The track points for the post 1944 data at six-hour time steps are binned into boxes, and the count in each box is recorded. This number was divided by the total number of years of data (68) to compute mean annual values. The observations are heavily clustered along the curve of the parabolic sweep with a maximum off the coast of the southern US (Fig 12.4a). Storm starting locations, being of particular interest, were also binned and plotted (Fig 12.4b). Births are clustered primarily in four regions covering approximately eight boxes, each having as many as 25 births over the 68 seasons.

Fig 12.5 shows the simulated corresponding figure of mean six-hour periods (Fig 12.5a) and births (Fig 12.5b). six-hour periods are slightly over estimated in the mean below 24° N and under estimated above that latitude. Exiting the tropics and entering the extra tropics cyclones are subjected to strong wind speeds and the six-hour observations are father apart. In an alternate run (not shown) a 5° radius is employed rather than the 2.5°. This change decreases the under estimation above 24° N, however it greatly increases the over estimation below that latitude. Mean simulated starts (Fig 12.5b), slightly overestimate the hurricane births off the coast of Africa which maybe the cause of the slight increase number of six-hour periods found there. Since these are randomly sampled unconditionally from the historical set the difference is purely due to sampling variations.

10.2 Comparing percentiles of the residence time in each grid box

The spatial structure of the simulated tracks was assessed through a comparison of the number of six-hour time steps in each box relative to the historical data. Since we have 1000 simulations from HITS, we compute the percentile of the historical residence time relative to these simulations as the basis for comparison. In Fig. 12.6 the color blue

indicates that the observations fall in the upper fifth percentile and green the upper quartile of the mean. The figure shows that the simulation tends to underestimate the number of six-hour periods in data sparse areas.

A comparison of number of six-hour periods is offered in Fig 12.7 for locations where mainland landfall is possible. A smoothed split violin plot (Fig 12.7) shows the historical (blue, left) and simulated (red, right) kernel density plots for number of the six-hour time periods. In addition, the 25^{th}, 50^{th}, and 75^{th} percentiles are marked with a line, and the mean values for historical (left) and simulated (right) are printed along the x-axis. Labeling of the x-axis location is not inclusive, but gives a geographical marker in the box considered. The overall positive skew of the distribution of six-hour time periods is captured by the simulation with smoothing as expected from the larger sample size of the simulations. Also the extreme tail of the data (red thin lines extending upward from main portion of the data) is typically better populated as expected from the simulations. Boxes in which six-hour time periods are more prevalent (Florida, South Carolina) show pockets of the blue observed values on the red long thin simulated extreme tails. HITS allows computation of events with potential return periods of 1000 years from the 1000 simulated years. This is in evidence from the extension of the tail relative to the observations.

The historical distribution is occasionally multimodal while the simulated is uni-modal reflecting the smoothing from the larger sample size. Differences in the underlying probability distributions of the historical and simulated are tested using a two-sample Kolmogorov-Smirnov (KS) and Cramer-von Mises (CM) tests. Both the KS and CM tests are nonparametric and compare the location and shape of the empirical cumulative

distribution functions of the two samples. All but one of the boxes (Virginia) pass the KS and CM tests at the 5% significance level indicating the same underlying probability distributions. Mean values of historical and simulated distributions on the y-axis are similar, along with the distribution shape. However, the median historical values fall between the 25th and 75th percentile of the HITS simulations in all except the low populated Texas box.

10.3 Landfall statistics

A land/sea mask was created to indicate which boxes in the domain are considered mainland, and which ocean. From that it was possible to count the storms (both historical and simulated) as they crossed from ocean to the mainland, and vice versa. A smoothed split violin plot of historical (blue, left) and simulated (red, right) distributions of mainland landfalls is shown in Fig 12.8. Width of the histogram is normalized to the maximum width equal to 0.9, the 25th, 50th, and 75th percentiles are marked with a line, and the mean values for historical (left) and simulated (right) are printed along the x-axis. Distributions of landfalls are remarkably similar between historical and simulated. In a majority of the boxes the simulated extreme tail extends beyond the historical (Belize, Yucatan, Mexico, Louisiana, Florida, S. Carolina, and Connecticut). The remaining landfall boxes are matched by the historical extreme value except Texas; which is underestimated. The Texas box appears in Fig 12.6 as an area in which low density of historical tracks leads to an underestimation of simulated tracks in those regions (median of the historical is above the 75th percentile of the simulation). As in the six-hours per year violin (Fig 12.7), all but one of the boxes (Virginia) pass the KS and CM tests at the

5% significance level indicating that the underlying probability distributions of the historical and simulated are the same.

10.4 Comparing frequency of six-hour periods of hurricane strength wind (>64 knots)

For six-hour time periods of hurricane strength (33 m s^{-1}, 64 knots, 74 mph) and above, a smoothed split violin plot (Fig 12.9) shows the historical (blue, left) and simulated (red, right) distributions in mainland landfall areas. The hurricane strength statistic is a subset of the six-hours per year (Fig 12.7), but the distribution of counts is different for the two: smoother and shorter simulation tails. The historical and simulated means are closer in value for the hurricane strength (Fig 12.9) than the six-hours per year (Fig 12.7), although the overall numbers of occurrences are reduced. All boxes pass the KS and CM tests at the 5% significance level indicating that the underlying probability distributions of the historical and simulated are the same; an improvement over the six-hours per year.

10.5 2012 Hurricanes

Since HITS was fit using the HURDAT dataset of 1851-2011, an out of sample of HITS on 2012 tracks was made by selecting different positions on a hurricane track to simulate from. This is a different way to apply the same model. Instead of looking at long terms simulations, one looks at a given hurricane track at a particular stage, and using the conditional distribution of trajectory and intensity given a position on the track, generate forward simulations. This was applied to several historical hurricanes with similar

results. Here, we present the comparisons for the recent hurricanes, Isaac and Sandy from 2012. In each case, different starting positions for the conditional simulation were considered prior to landfall, and 1000 simulations were performed for each starting position. These were compared with the NOAA hurricane forecasts from the same locations.

Text files of latitude, longitude, and wind speed measurements every six-hours, as in the dataset HURDAT, were taken from the Atlantic Hurricane Track Map & Images page of Johns Hopkins University (http://fermi.jhuapl.edu/hurr/index.html). The 5-day track forecast, uncertainty cone, and watch/warning images were obtained from National Weather Service National Hurricane Center and near surface daily wind speed over land and ocean from the NCEP-National Center for Atmospheric Research (NCAR) Reanalysis (Kistler et al. 2001) 0.9950 sigma level.

10.5.1 Sandy

Hurricane Sandy was a devastating storm in 2012, making landfall in Jamaica as a category 1 storm on the Saffir-Simpson Hurricane wind scale, in Cuba as a category 3, and in southern NJ as a post-tropical cyclone with a significant storm surge in the mid-Atlantic and New England states (Blake et al. 2013). Fig 12.10 is the HITS hurricane wind strength area in six-hour time periods per year (a), maximum sustained winds from 26

October 2012 to 31 October 2012 (b), simulated mainland landfalls (c), and watch/warning image from National Weather Service National Hurricane Center (d) for 26 Oct 2012 at 5 PM. EDT. Black circles on columns one to three mark the actual path

of Hurricane Sandy and a large black x indicates where the simulation was started. On 26 October, three days before landfall in southern NJ, Sandy is off the Florida coast with HITS showing high values of hurricane strength six-hour time periods and mainland landfalls along the Eastern coast. The actual path of Sandy in Fig. 12.10a is within the HITS hurricane strength wind area over the ocean. The solid contour of Fig. 12.10b indicates Sandy's maximum sustained hurricane strength winds and is similar to the size and shape of the simulated hurricane wind area in Fig. 12.10a. HITS mainland landfalls in Fig. 12.10c have bulls-eyes on Florida and the Mid-Atlantic. On 27 October (not shown) HITS no longer has the maximum landfall in Florida, as the hurricane is farther north and east by then. The NHC forecast, in Fig. 12.10d shows landfall in Delaware, south of the eventual landfall location.

10.5.2 Isaac

Hurricane Isaac passed over the Lesser Antilles, Haiti, and eastern Cuba as a tropical storm, and intensified to a category 1 hurricane before making landfall in southeastern Louisiana causing storm surge and inland flooding to southeastern Louisiana and southern Mississippi (Berg 2013). Fig 12.11 is the HITS hurricane wind strength area 6 hour time periods per year (a), maximum sustained winds from 25 Aug 2012 to 30 Aug 2012 (b), simulated mainland landfalls (c), and watch/warning image from National Weather Service National Hurricane Center (d) for 25 Aug 2012 at 5 AM EDT. Black circles on columns one to three mark the actual path of Hurricane Isaac and a large black x indicates where the simulation was started.

On 25 August, Isaac is over Cuba and the actual path in Fig. 12.11a is within the HITS hurricane strength wind area over the ocean. The solid contour of Fig. 12.11b indicates Isaacs's maximum sustained hurricane strength winds and is similar to the size and shape of the simulated hurricane wind area in Fig. 12.11a. HITS mainland landfalls in Fig. 12.11c have a large value over Florida with smaller ones on the Gulf coast, East coast, and even Central America. On 27 August (not shown) HITS has a greater probability of landfall on the Gulf coast as it moves into the Gulf of Mexico. The NHC forecast, in Fig. 12.11d shows landfall in western Florida, east of the actual Louisiana landfall.

This page intentionally left blank

Chapter 11

Summary and Discussion

A new hurricane track simulator that is motivated by the observation through initial analyses that Markovian models tend to be more diffusive that is apparent from the historical observations was presented. The basic idea considers conditional distributions along a track that are empirically estimated using nonparametric density estimation applied to the historical data and selected track attributes. The algorithm bears some similarities to a Markov Renewal model, in that it considers the time to spend on a track (state) conditional on selected attributes as the state variable rather than a discrete transition from one grid box to another, or a march to a displaced location at each time step. In fact, for model development space is considered continuous, with no gridding, while time is discrete following six-hour steps. The local tracks at a point are considered candidate states, and a transition to one of them is selected based on distance between tracks, track orientation and the age of each of the tracks. The last variable allows a consideration of the past history of each of the tracks and hence discriminates between tracks that may have been just born near that location vs. ones that were born quite a bit further. Thus, the length of time in a past state is accounted for and the length of time to spend in the new state is simulated, similar to a Markov Renewal model. The result is an

improvement in the residence time, spatial distribution and landfall densities of the simulations over a Markov chain based model for track simulation.

Performance of the HITS simulated tracks was evaluated over 5º by 5º boxes for a number of statistics. While some biases are evident in the areas poorly sampled in the historical data base, it is clear that the simulations preserve the essential attributes of residence time, spatial patterns and landfall, especially for the stronger wind thresholds: the six-hour periods of hurricane strength winds for all mainland landfall boxes passed KS and CM tests at the 5% significance level indicating that the historical and simulated were from the same distribution. Even though the model is based directly on the historical record and is nonparametric, an extension of the tail probability distribution and smoothing of the probability distribution of the statistics of interest is seen relative to the historical data. Similarly, conditional simulations of historical tracks showed propagation dynamics that relative to the point they are started from have performance similar to those produced by dynamical models that are in use for near real time hurricane forecasts. We do not suggest HITS as a forecast model since none of the essential physics is modeled at all. However, it seems that the information contained in the historical tracks does contain enough of the location relevant physics such that the model that simulates tracks based on geometrical similarity criteria is able to do a conditional simulation of the tracks from different locations.

Several track simulation modelers have divided up the Gulf and U.S. coast and compared their model landfall results to the historical HURDAT (Figs 4 and 5 and table 1 of Halegate 2007, Fig 18 Hall and Jewson 2007, Fig 3 of Vickery 2000, and table 1 in

Rumpt 2008). The areas are different though for all models, so they cannot be compared directly. They all use different statistical tools to judge the "goodness" of the fit.
Hall and Jewson (2007) also use the number of six hourly tropical cyclone positions per area in their Fig 15. Emulating their Fig 15d "Z score" (normalized probability of historical minus simulated mean divided by historical), all of the HITS boxes were between -1 and +1 except for those where historical and simulated counts were zero leaving an undefined value. The mean normalized probability was 0.063. Although it is important to compare the mean (or median) of simulated and historical data this analysis emphasized comparing the shape of the entire distribution as extreme events appear on the tail end. Research has shown not only do tropical cyclone distributions display a heavy tail (Figs 12.7-12.9), but also hurricane damage is heavily tailed (Katz 2002).
In sensitivity testing of the HITS model:

- A 5º degree radius of neighbor points was tried rather than the 2.5º (given as D).

This larger radius sampled among unlike populations of tracks in the tropics and was abandoned.

- Several ways of computing the vector were attempted: with previous, current, or future points, as vector differences, or as distance angle vectors.

Computing both angles and distances with future points (distance or angle needed to jump to next track segment) gave the best results.

- Median of track lifetimes was used rather than random draw (given as L).

Simulated track length was unrealistic using median, as shorter and longer tracks were not represented.

- Use of only post 1944 data.

The best results came from using all available data. If quality of data is a concern selection of those tracks can be down weighted. Behavior of a system with determinism like the paths of N. Atlantic tropical cyclones is best studied with all available information on past behavior.

There are very few parameters of the HITS model that are specified by the user; the model is data driven. For the analysis, grids of one, three, five, and ten degree "boxes" were tested. Smaller boxes showed fewer number (compared to the total) of boxes in lower and upper fifth percentiles, but less data per box gives less reliable statistics. Other changes were postulated, but not simulated for example time/point step changes. The time step of six hours could be increased to 12 or even 24-hour steps. We looked at probabilities of moving to one point on a track at a time, but whole track segments could be included.

Currently, our future work plan includes running the model backwards to determine where all land falling storms in a particular box started. We also plan to explicitly consider conditioning on large-scale climate variables to see if inter-annual variability in tropical cyclone counts and tracks can be properly simulated. As clustering results of N. Atlantic tropical cyclone tracks have shown groupings that display differing genesis locations, track shapes, intensities, life spans, landfalls, seasonal patterns, and trends

(Nakamura et al. 2009), selecting the birth location based on climate state would also impact the resulting tracks and probabilities.

This page intentionally left blank

Chapter 12

Figures

Figure 12.1: 1944-1999 six-hour per year periods of cyclones starting in the box 10N to 15N and 45W to 50W historical (top) and Markov chain model (bottom).

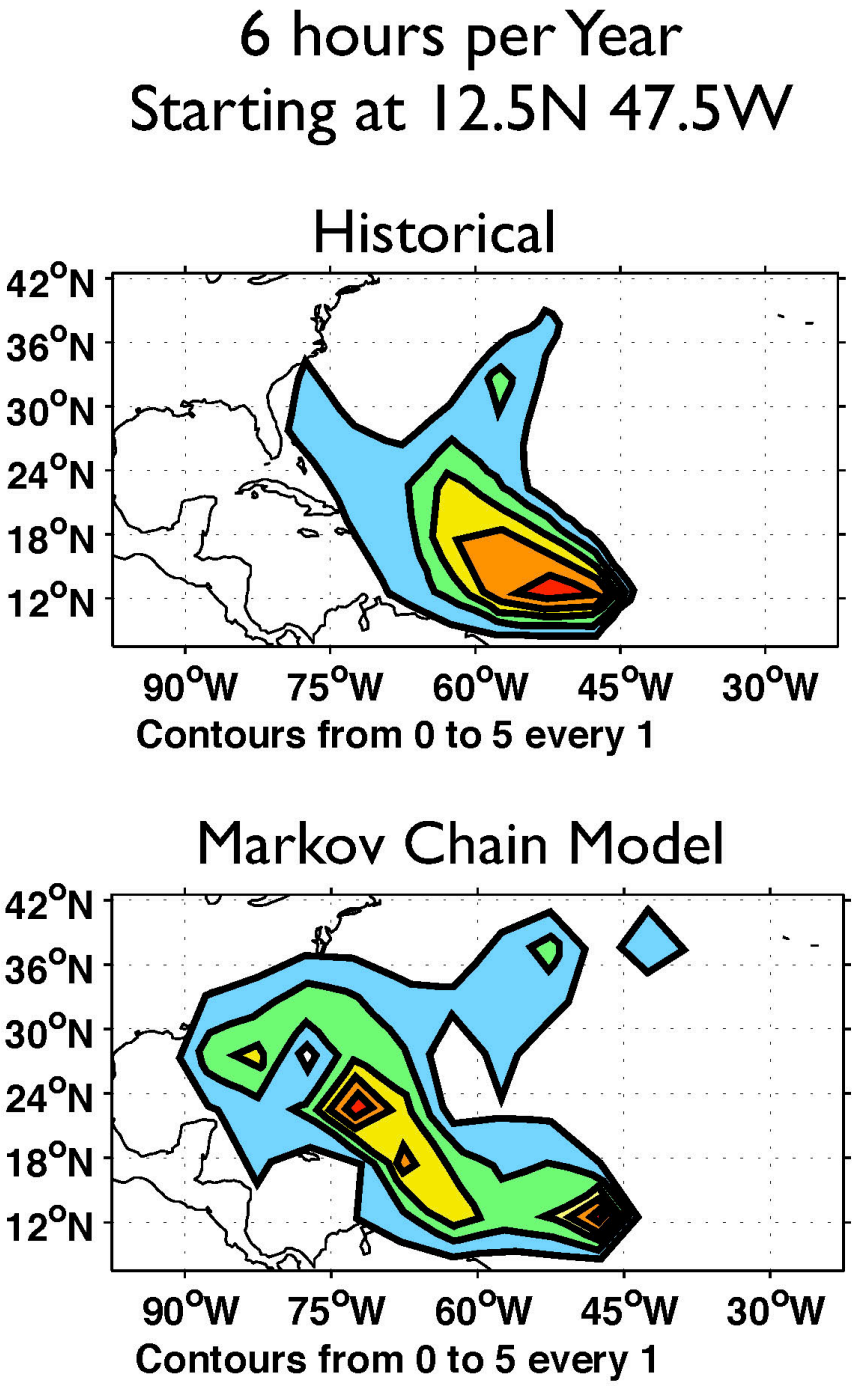


Figure 12.2: Abbreviated flow chart of HITS algorithm: an illustration of steps.

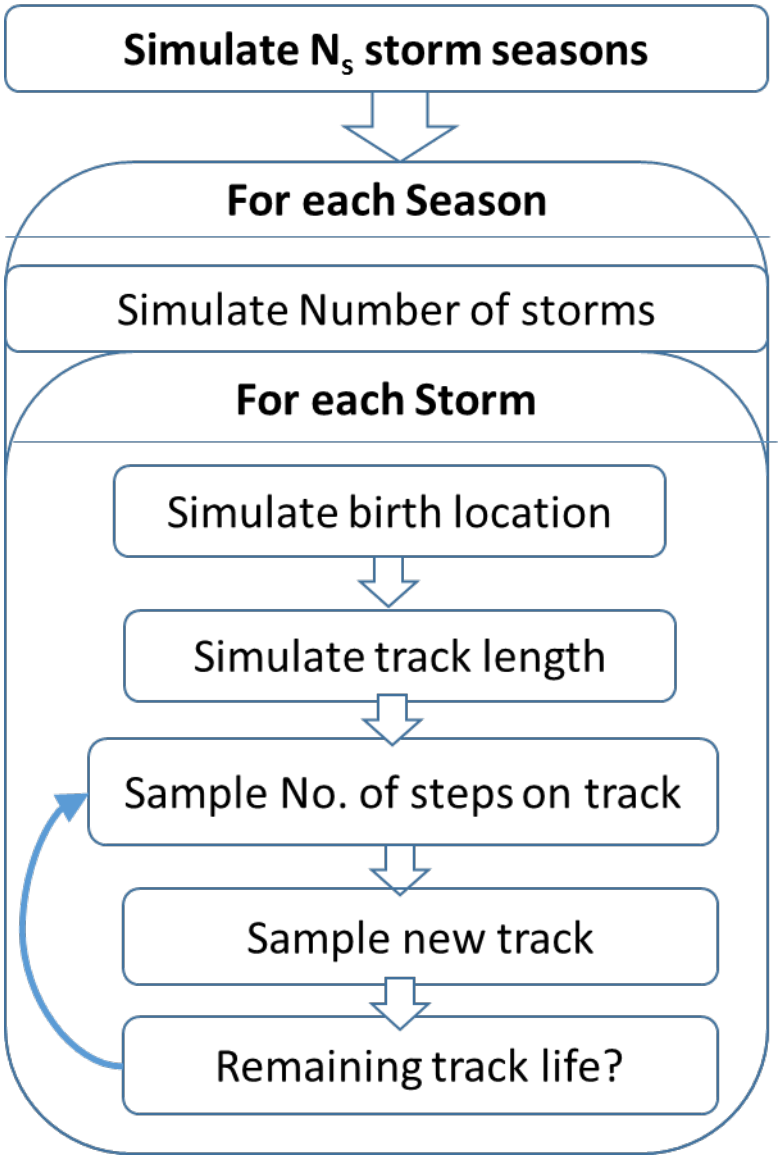


Figure 12.3: The 2011 North Atlantic hurricane season tracks (top) and a simulated North Atlantic hurricane season using the hurricane interactive track simulator (bottom).

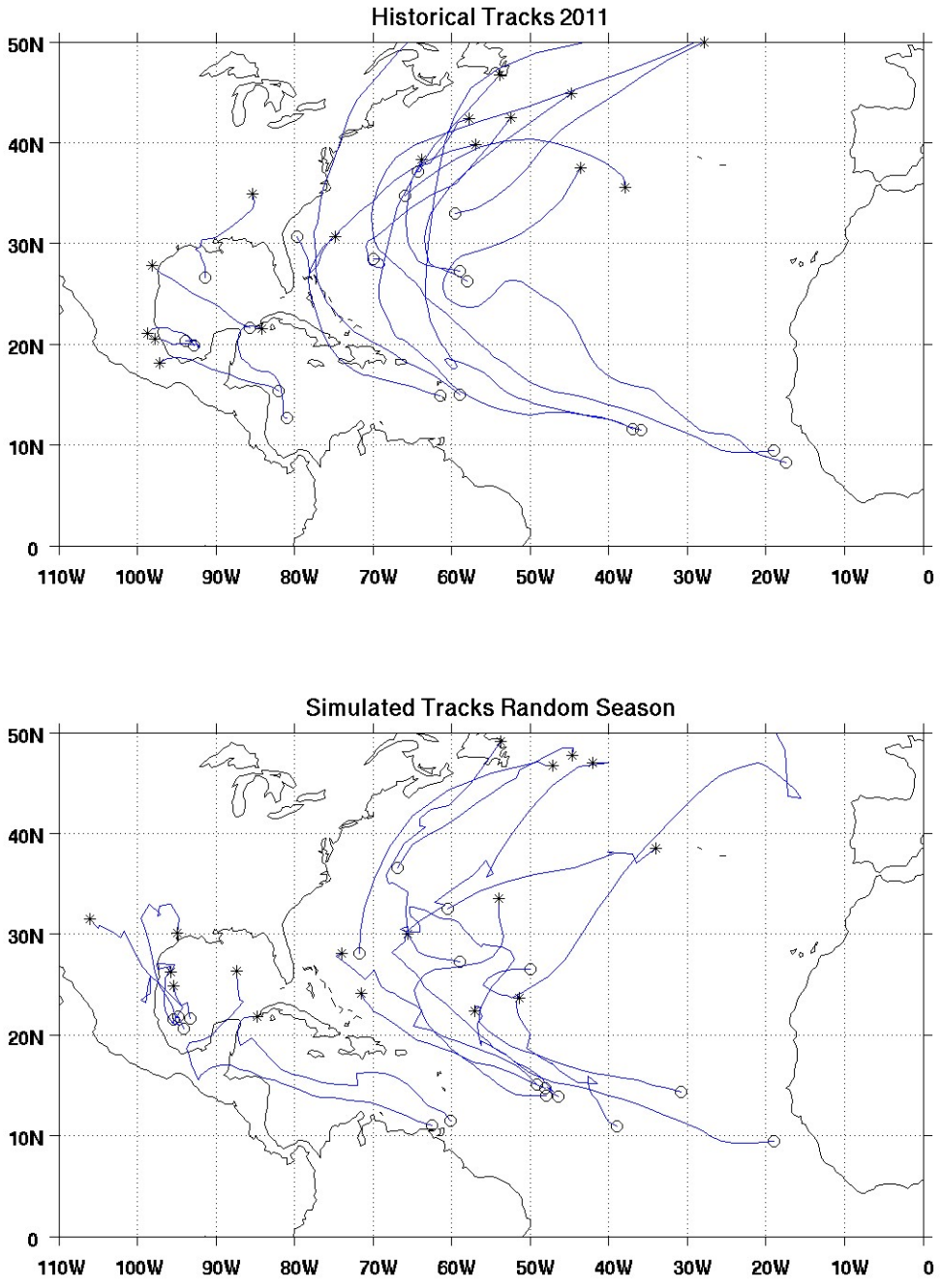


Figure 12.4: Average number of historical (1944-2011) tropical cyclone six-hours per year periods binned into 5° by 5° degree grid boxes (top) and number of historical tropical cyclone starts per box (bottom).

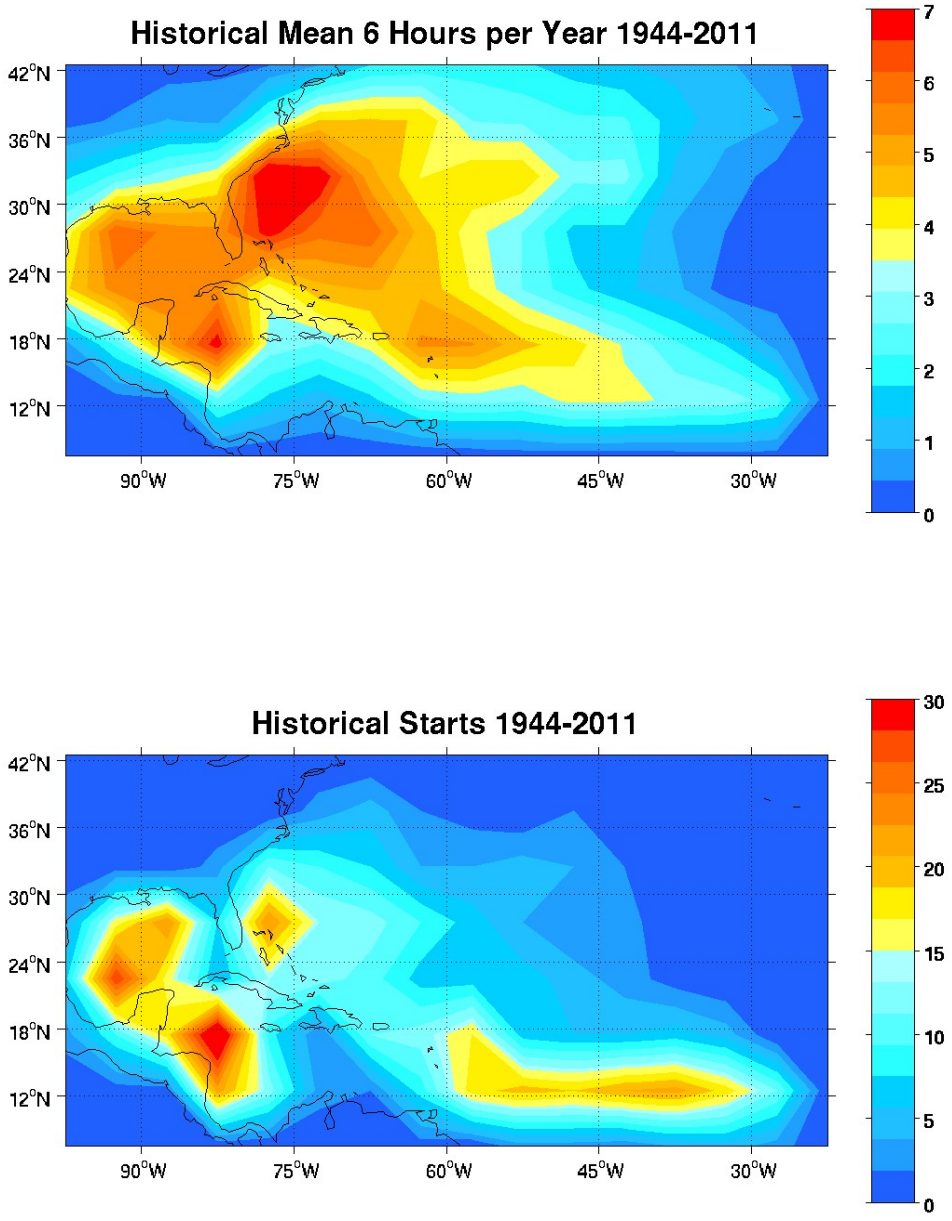


Figure 12.5: Average number of simulated tropical cyclone six-hours per year periods binned into 5° by 5° degree grid boxes (top) and number of historical tropical cyclone starts per box times 68 to match historical record of 1944-2011 (bottom).

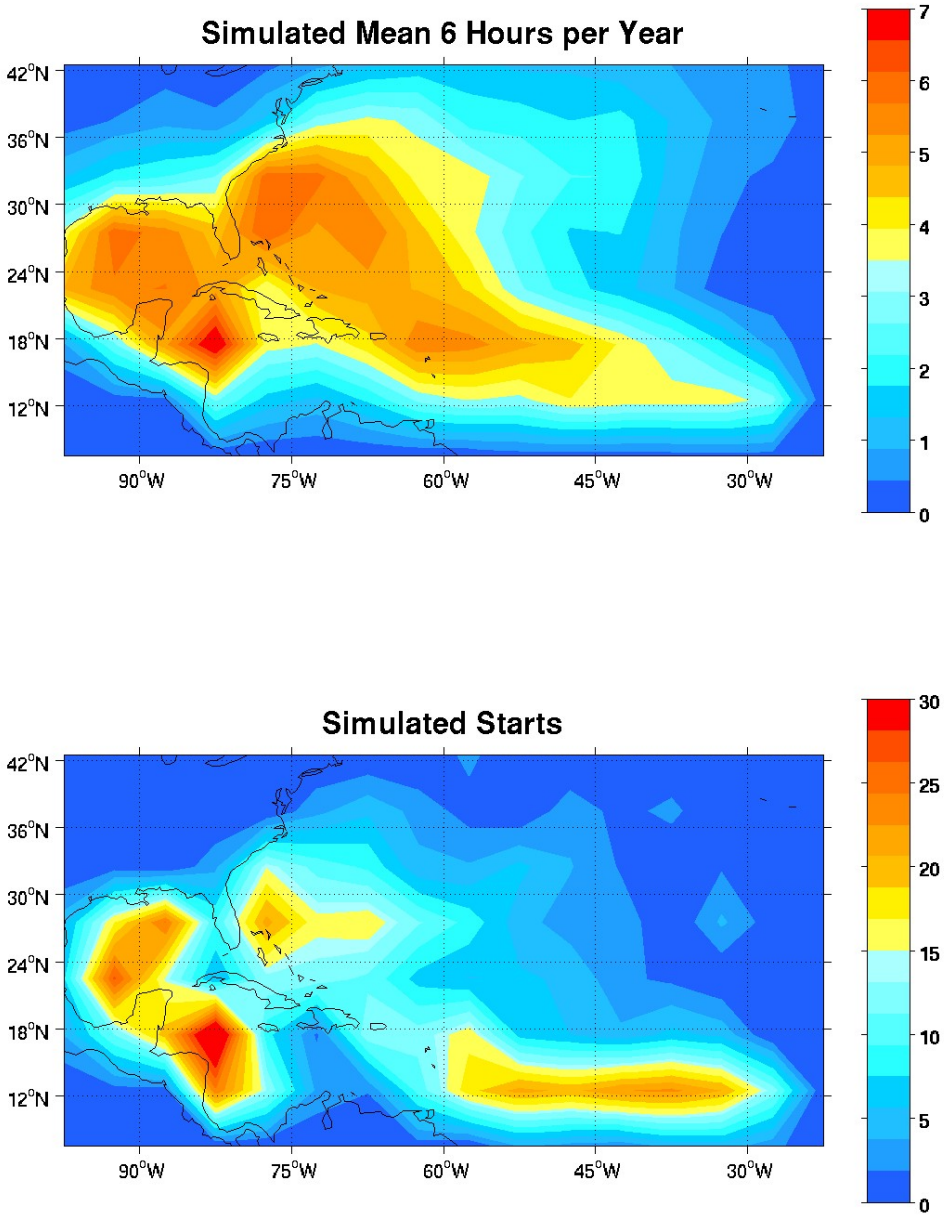


Figure 12.6: The spatial structure of the 1000 simulated tracks shown by the percentiles of the count of six-hour time steps in each box compared to historical (1944-2011). The color blue implies that the observations fall in the upper fifth percentile and green the upper quartile.

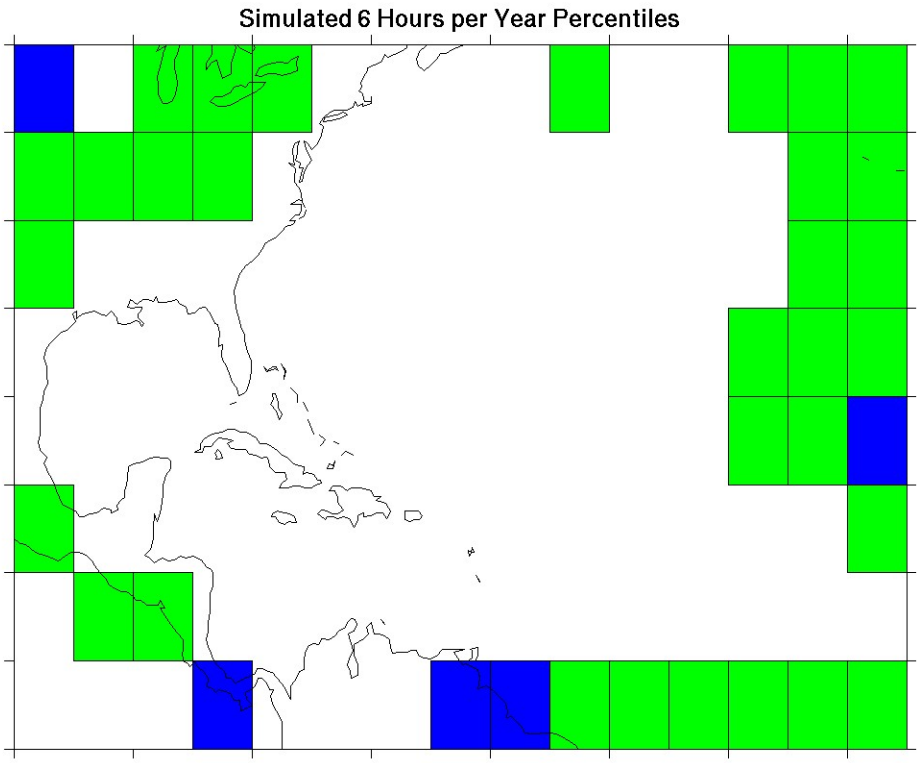


Figure 12.7: Kernel density estimation of observed 1944-2011 (blue, left) and simulated (red, right) six-hours per year split violin plot for landfall areas. Mean values for observed (left) and simulated (right) are located on the x-axis.

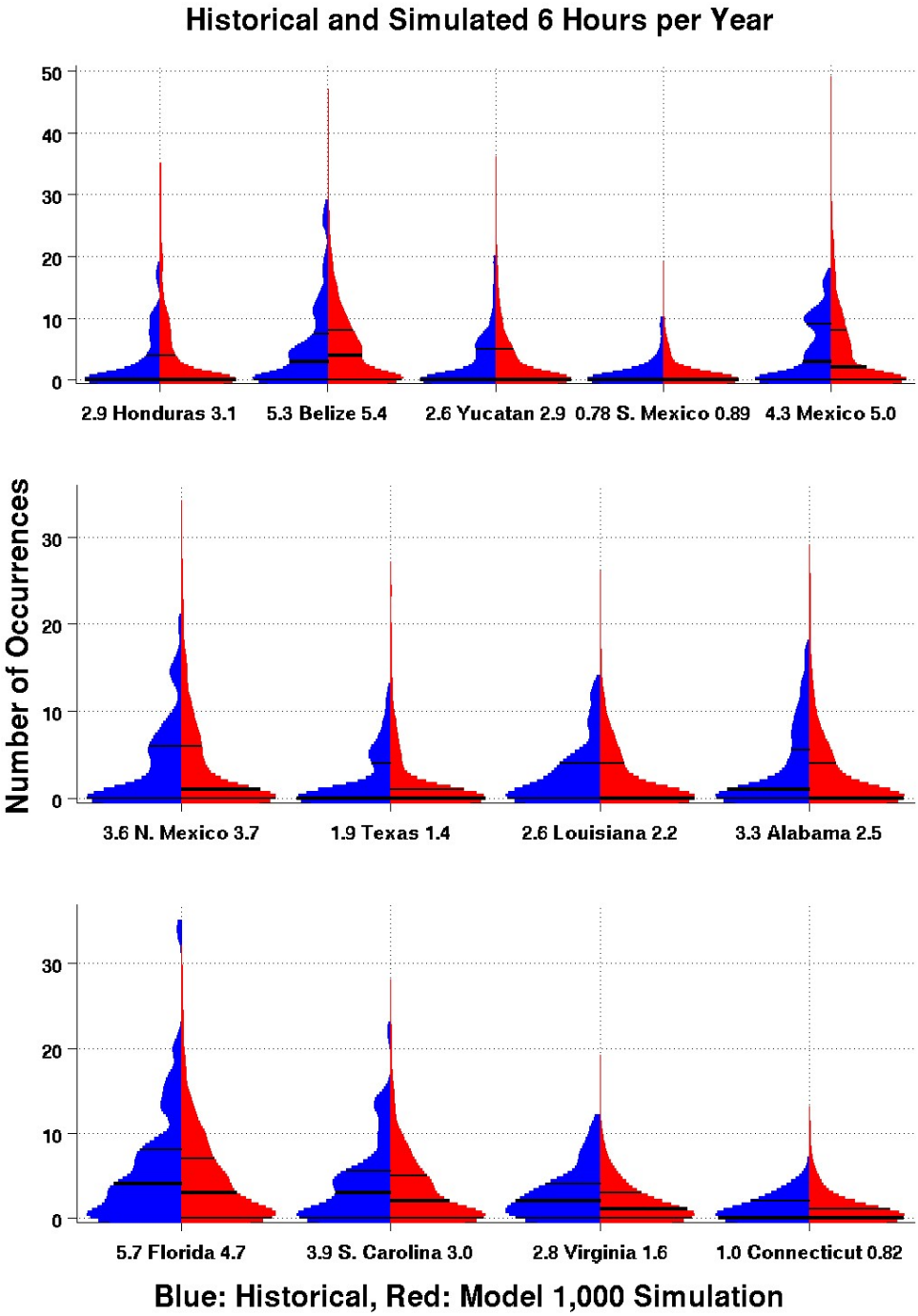


Figure 12.8: Observed 1944-2011 (blue, left) and simulated (red, right) split violin plot of number of tropical cyclone landfalls per year for landfall areas. Mean values for observed (left) and simulated (right) are located on the x-axis.

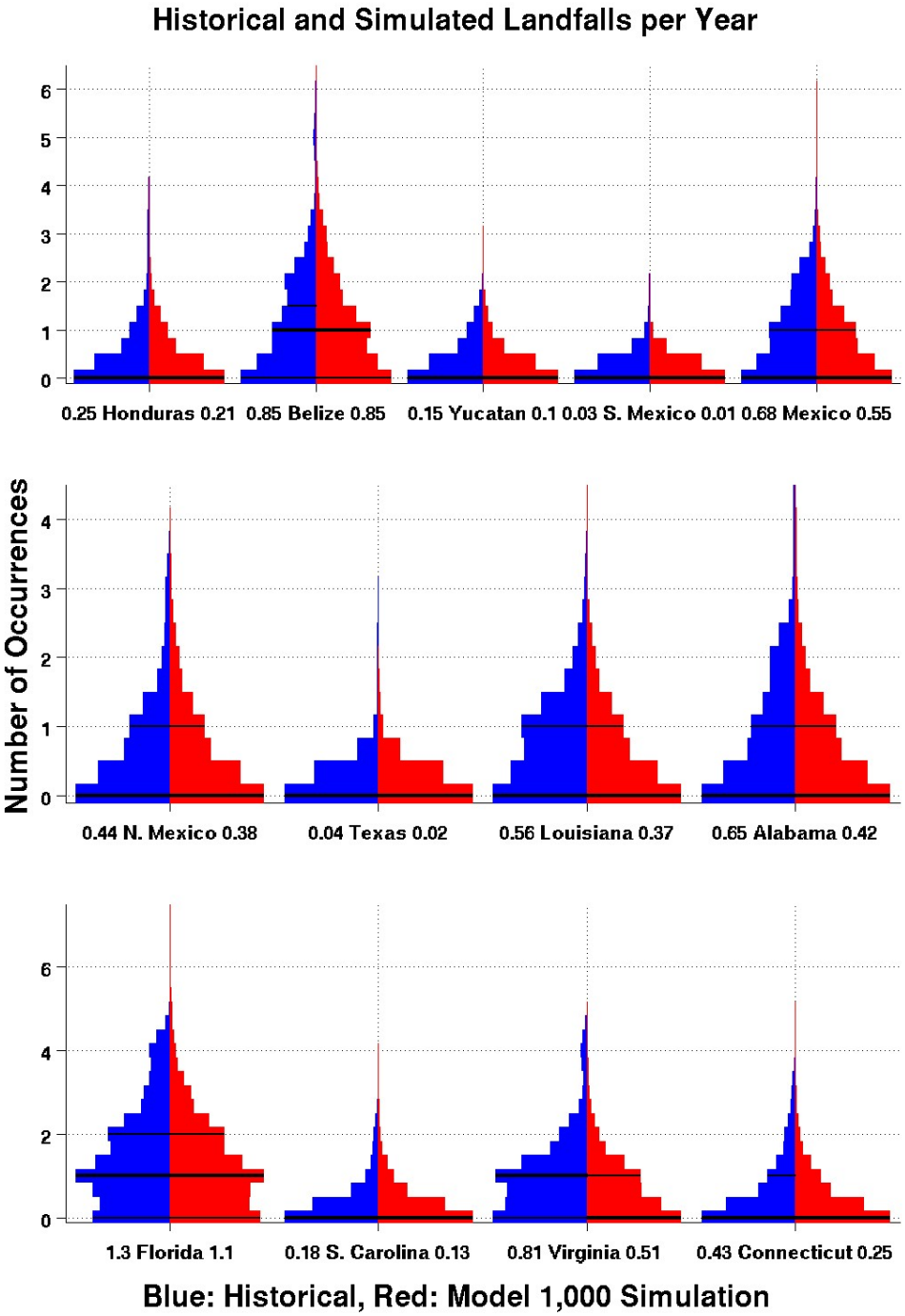


Figure 12.9: Kernel density estimation of observed 1944-2011 (blue, left) and simulated (red, right) of hurricane strength winds in landfall areas. Mean values for observed (left) and simulated (right) are located on the x-axis.

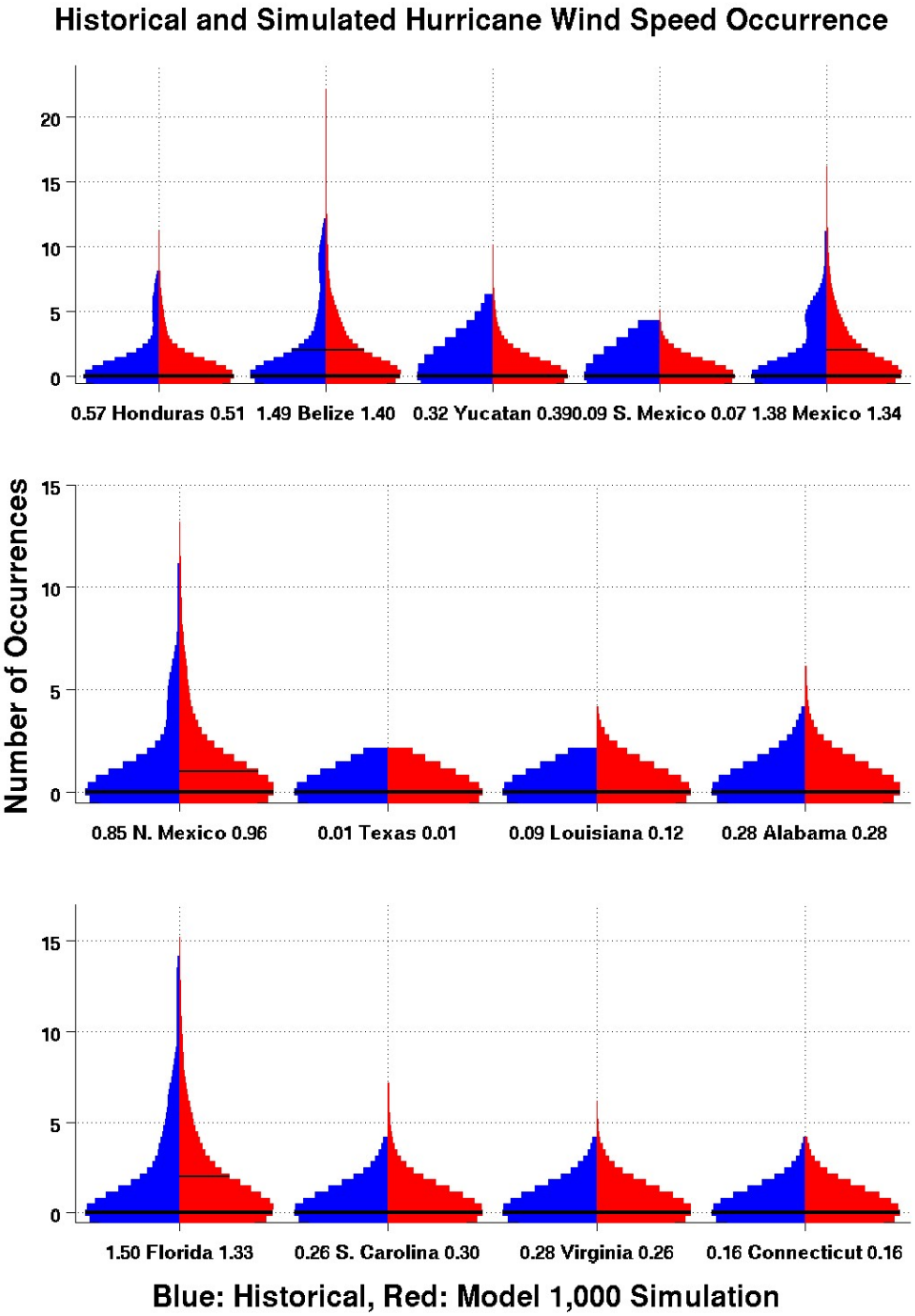


Figure 12.10: Hurricane Sandy, simulated hurricane wind strength area in six-hour time periods per year (a), maximum sustained winds from 26 Oct 2012 to 31 Oct 2012 (b), simulated mainland landfalls (c), and watch/warning image from National Weather Service National Hurricane Center (d) for 26 Oct 2012 at 5 PM EDT. Black circles on a, b, and c indicate the actual path of Hurricane Sandy and a large black x indicates where the simulation was started.

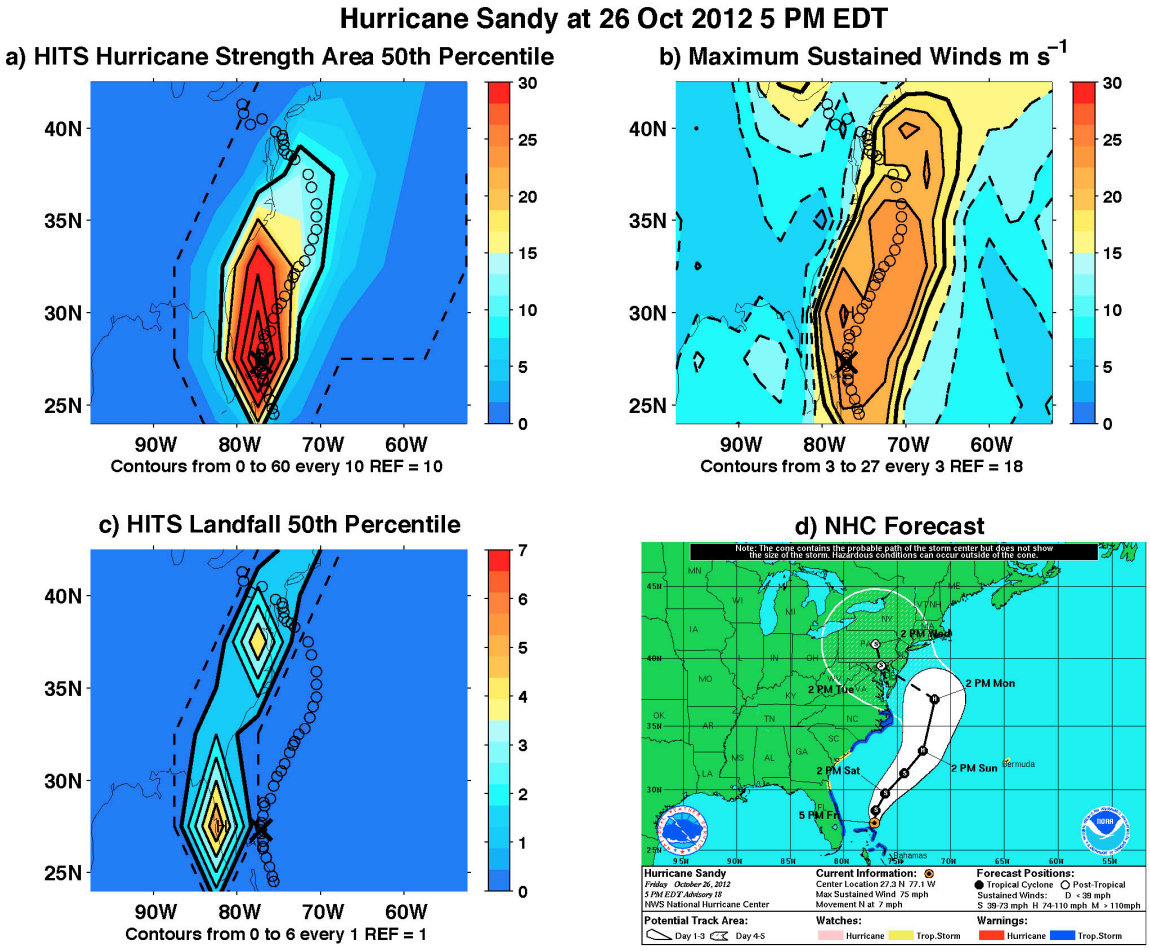
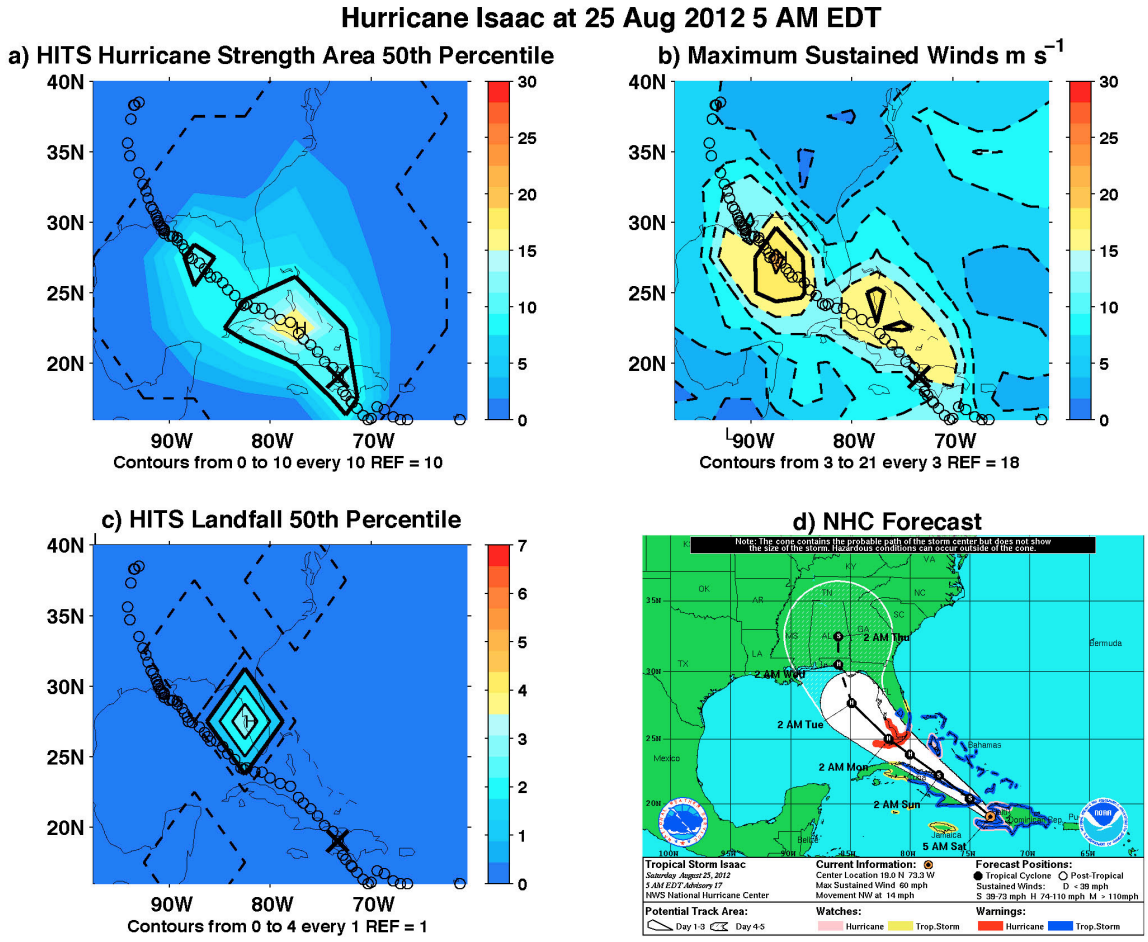


Figure 12.11: Hurricane Isaac, simulated hurricane wind strength area six-hour time periods per year (a), maximum sustained winds from 25 Aug 2012 to 30 Aug 2012 (b), simulated mainland landfalls (c), and watch/warning image from National Weather Service National Hurricane Center (d) for 25 Aug 2012 at 5 AM EDT. Black circles on a, b, and c indicate the actual path of Hurricane Isaac and a large black x indicates where the simulation was started.



This page intentionally left blank

Part III

Dynamical Structure of Extreme Floods in the U.S. Midwest and the United Kingdom.

This page intentionally left blank

Twenty extreme spring floods that occurred in the Ohio Basin between 1901 and 2008, identified from daily river discharge data, are investigated and compared to the April 2011 Ohio River flood event. Composites of synoptic fields for the flood events show that all these floods are associated with a similar pattern of sustained advection of low-level moisture and warm air from the tropical Atlantic Ocean and the Gulf of Mexico. The typical flow conditions are governed by an anomalous semi-stationary ridge situated east of the US East Coast, which steers the moisture and converges it into the Ohio Valley. Significantly, the moisture path common to all the 20 cases studied here as well as the case of April 2011 is distinctly different from the normal path of Atlantic moisture during spring, which occurs further west. It is shown further that the Ohio basin moisture convergence responsible for the floods is caused primarily by the atmospheric circulation anomaly advecting the climatological mean moisture field. Transport and related convergence due to the covariance between moisture anomalies and circulation anomalies are of secondary but non-negligible importance. The importance of atmospheric circulation anomalies to floods is confirmed by conducting a similar analysis for a series of winter floods on the River Eden in northwest England.

Citation: Nakamura, J., U. Lall, Y. Kushnir, A.W. Robertson and R. Seager, 2013: Dynamical Structure of Extreme Floods in the U.S. Midwest and the United Kingdom. *J. Hydrometeor.*, **14**: 485-504.

With the addition of section 15.2 Dipole Index and Extreme Precipitation Relation and the accompanying figures 17.12 – 17.14.

Chapter 13

Introduction to Part III

Floods account for a large fraction of losses due to natural hazards (FIFMTF 1992). The combination of large air thermal contrasts, larger amounts of water vapor in warmer air, saturated soils, frozen ground, and snowmelt make spring the season with the greatest flooding potential in most of the United States (Hirschboeck 1991). The subject of our study, the Ohio River, is the largest tributary, by volume, of the Mississippi River. The March 1913 flood in the Ohio River Basin tops the list of significant regional floods in the United States in the 20th century in terms of number of lives lost and property damage (Perry 2000). The April 2011 flood event that contributed to the subsequent lower Mississippi River floods was another major event and ranks in the top 5 spring events in the 85-year record at the East Fork of the White River in Shoals, IN, a tributary of the Ohio River.

It has been recognized that, especially for larger river basins, a link between an oceanic moisture source and local precipitation is likely necessary for generating the extreme precipitation associated with a large regional flood (Hirschboeck 1991). Antecedent soil moisture or snow melt may also contribute to enhanced flood potential but may in turn be the consequence of prior moisture transport into the basin from the same oceanic source.

Thus, for extreme floods, the large-scale meteorological conditions associated with organized transport of moisture from oceanic sources to a continental region are likely to be important. The atmospheric flow anomalies associated with some major floods have been shown to be predictable to some extent in the weather forecast context. Whether they are predictable, in a probabilistic sense, on longer, seasonal, timescales, using either dynamical or statistical models, is not yet clear. Such an assessment requires an understanding of the underlying physics, which is what we attempt here.

Hirschboeck (1991) laid out a classification for the climatic context of floods, identifying the key seasons and the general pathways of atmospheric moisture transport associated with large regional floods. However, to date there has been only a limited quantitative application of her ideas to the problem of understanding the recurrence of floods and predicting when and where they will occur. The concept of 'atmospheric rivers' (Zhu and Newell 1994, 1998), and their connection to flood events (Ralph et al. 2006; Neiman et al. 2011) is a notable advance in this direction. The term atmospheric river (AR), coined by Zhu and Newell (1994), refers to the long (planetary scale) filaments of atmospheric moisture seen in the vertically integrated water vapor field and that are associated with fast moving atmospheric flows. These atmospheric features are transient and not spatially fixed and can be best seen in daily satellite picture and weather analyses (Zhu and Newell 1998). Much of the analysis associated with these phenomena was focused on Pacific Ocean cases that carry moisture to the western US, the so-called "Pineapple Express" because of the link back to the Hawaii region of the subtropical Pacific (Dettinger 2011; Dettinger et al. 2011). Dirmeyer and Kinter (2009) identified similar features carrying moisture from the tropical Atlantic into the U.S. Midwest in May-July and coined the

term 'Maya Express'. They found that this feature is due to a strengthening or westward shift of the Atlantic subtropical ridge. These authors also state that, large scale flooding over the U.S. Midwest is not local, but part of a large-scale circulation change connecting the tropics and mid-latitudes (Dirmeyer and Kinter 2010). Moore et al. (2012) examined a case study in which ARs from both the eastern tropical Pacific and the Caribbean Sea lead to the formation of mesoscale convective systems and subsequent flooding in Tennessee and Kentucky. In a recent analysis focused on the Atlantic Basin, Lavers et al. (2011) showed ARs to be present at the top 10 flood events since 1970 for several river basins in Britain.

While much flood related research has focused on anomalous moisture transports it is also the case that the climatological moisture transport occurs in specific regions. The climatology of tropical moisture export (TME) assembled by Knippertz and Wernli (2010), provides a framework for understanding the pattern of large moisture advection paths out of the northern hemisphere tropics. North American tropical flows in the spring include the Great Plains (GP) TME between the Rockies and the Appalachians, and the Gulf Stream (GS) TME to the east of the Appalachians. In spring at 35°N Knippertz and Wernli (2010) found the GP and GS TME well separated, both concentrated predominately below 600 mb, with the strongest GP TME at approximately 860 mb and the middle of the GS TME higher at about 780 mb. In winter at 50°N, GS trajectories show strong ascents into the mid-troposphere and the tracks shift eastward by as much as 20° longitude (Knippertz and Wernli 2010). At this latitude band, in winter, GS TME can reach the Northern European coast and Great Britain.

In this paper we examine the meteorological context of major Ohio River Basin floods (identified as a 10-year return period in annual maxima of stream flow) and examine the conditions that led to the flooding in April of 2011 in this context. We demonstrate that these floods are associated with a recurrent type of persistent, anomalous large-scale circulation pattern that is responsible for conditions that lead to atmospheric moisture flows into, and convergence over, the basin and heavy precipitation during each of the 20 large spring (March-to-May) floods identified between 1901 and 2008. We then show that the April 2011 Ohio River flood exhibited very similar meteorological conditions and multi-day evolution as the composite of the previous 20 major floods, thus confirming the unique meteorological state that leads to extreme floods in the Ohio River Basin.

We further demonstrate that the apparent ARs and floods are caused primarily by the circulation anomalies advecting and converging the background mean, climatological moisture field. This is also found to be the case in the very different meteorological environment of winter floods in Great Britain (Lavers et al. 2011). Both case studies illustrate the importance of atmospheric circulation variability on daily to seasonal timescales in causing extreme floods.

This page intentionally left blank

Chapter 14

Data

Twenty Ohio basin floods that occurred prior to the April 2011 event were selected by analyzing daily river discharge data from 7 gauging stations associated with sub-basins of the Ohio River larger than 10^3 km^2. The events and stations are listed in Tables 17.1 and 17.2. These 7 stations were originally identified as part of the Hydro-Climatic Data Network (HCDN) of the U.S. Geological Survey (USGS), and their data are reportedly free of regulation and diversion. For each station, events where the peak flow exceeded the 10-year flood during the spring (March–May) season were identified for years 1901 to 2008.

The ten year flood event at each station was computed according to the U.S. Interagency Advisory Committee on Water Data publication "Guidelines for Determining Flood Flow Frequency", Bulletin 17-B of the Hydrology Subcommittee, with the addition of digitized generalized skew values extracted from the program PeakFQ (http://water.usgs.gov/software/PeakFQ). The assumption is that the flood information is a reliable representation of random homogeneous events. This includes climatic time invariance, floods as independent events, no watershed changes, similar types of events, and good data quality. Selection of HCDN stations by the USGS was made to meet as

many of the above assumptions as possible resulting in many fewer stations than the original USGS data.

Once the 10-year flood frequency was computed for every station using water year (October through September) annual peak maximum flow, daily data were processed to identify dates when the peak flow exceeded the 10-year threshold. Most flood events occurred at one station, but a few occurred simultaneously at several. The earliest flood date was selected for a multiple sub-basin flooding event. Some flooding events last a day and some weeks. In order to analyze pre-flood conditions, day 0 is defined as the first day of a possible series of flooding days.

Day 0 for comparison floods in Great Britain are the eight floods identified in Lavers et al. (2011) that occurred before 2008 on the River Eden at Temple Sowerby in north-west England. These floods were found by taking the top river flow dates from the UK National River Flow Archive of the winter half-year (October thru March) daily mean maximum. Temple Sowerby is the largest sub-basin in the Upper Eden catchment, with an area of 616.4 km^2, and the rainfall regime is predominantly frontal with a maximum in the winter (Ockenden and Chappell 2011).

The distribution of all flooding events for gauged Ohio River sub-basins greater than 10^3 km^2 shows they occur year round, with a median in February and a mean in mid-March. However, the distribution is heavily positively skewed towards the spring months with the seventy-fifth percentile at about 10 May. A third of the extreme floods occur in spring, the largest percent of any of the seasons.

Given that most of the floods exceeding the nominal 10 year exceedance level occurred in the first half of the 20th century, we use daily averaged 20th Century Reanalysis V2 data (20CR, Compo et al. 2011) as the principal source of atmospheric data for the historical analysis for both the Ohio and Eden basins. The original 20CR six hourly output is provided by the National Oceanic and Atmospheric Administration's Earth System Research Laboratory in Boulder, Colorado, USA (http://www.esrl.noaa.gov/psd/). The 20CR is derived by driving the April 2008 experimental version of the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) with observed sea surface temperatures boundary conditions and assimilating historical surface pressure observations only (Compo et al. 2011). The 20CR is a new product that is rapidly being recognized as a source of meteorological data with high temporal resolution, appropriate for diagnostic studies (Emanuel 2010; Compo et al. 2011; Truchelut and Hart 2011; Dole et al. 2011; Barriopedro et al. 2011). Lavers et al. (2011) used the 20CR in their prior study of floods in Great Britain including those on the River Eden at Temple Sowerby. We tested the robustness of the 20CR results by repeating our analysis using the standard NCEP-National Center for Atmospheric Research (NCAR) Reanalysis (Kistler et al. 2001) for those Ohio River floods that occurred after 1948. The results obtained (not shown) were essentially the same as those obtained with the 20CR.

Historical precipitation data are from the National Climatic Data Center's Global Historical Climatology Network (GHCN) daily station data (Vose 1992) gridded over land into one-degree boxes. The April 2011 event data fields are from the daily NCEP-NCAR Reanalysis (Kistler et al. 2001).

This page intentionally left blank

Chapter 15

Results

15.1 Ohio Basin Floods

Climatological, vertically integrated surface - 600 mb moisture flux and its convergence (expressed in mm day^{-1}) for spring illustrate the typical transport pattern and the normal sources and sinks of moisture, respectively (Fig. 17.1). Moisture sources, shown by low-level divergence of the moisture flux, are present in a broad region of the tropical Atlantic with smaller pockets in the Caribbean Sea and Gulf of Mexico. Vectors of moisture flux show transport of the moisture from the tropical Atlantic, through the Caribbean Sea, and turning northward through the Gulf of Mexico as the low-level flow encounters the lee-side trough on the eastern side of the Sierra Madre and Rocky Mountain chains. Moisture sinks, shown as low-level convergence, are present on the high plains just east of the mountains. This climatological pattern is in contrast to the anomalous moisture flux and convergence apparent during an extreme Ohio basin flood (see below).

To show the spatial extent of the precipitation patterns that lead to flooding of the Ohio basin the precipitation anomaly is plotted in Fig. 17.2 for the composite of the 20 historical Ohio basin flood events (panel a) and for the April 2011 event (panel b), each

averaged over the 9 days leading up to the floods together with the flood day (designated hereafter as days -9 to 0). Both have a positive anomaly over much of the eastern U.S. and extending into Canada indicating these are large-scale events. Weak negative anomalies are seen to the west, over the high plains at the foothills of the Rocky Mountains. The April 2011 event has stronger positive anomalies in the western portion of the basin and corresponding heavier flooding in that area, an expected result when a single event is compared with an average of 20 temporally independent events.

Fig. 17.3 shows the composite, vertically integrated surface - 600 mb daily moisture flux and its convergence for the 20 historical Ohio basin flood events (panel a) and for the April 2011 event (panel b), averaged over days -9 to 0 of all events, for both the full field (left) and the anomaly relative to a 1961-1990 climatology (right). For the total field (Fig. 17.3, left panels), for both the canonical historical events and the April 2011 event, there is a strikingly similar large-scale pattern of anti-cyclonic flow of moisture originating in regions of large-scale divergence in the tropical Atlantic, flowing across the Caribbean and Gulf of Mexico and turning northward into the Plains and northeastward to converging over the Ohio basin. The anomalous moisture flux (Fig. 17.3, right panels) is concentrated at the western flank of the Bermuda High showing southwesterly flow towards the Ohio basin from the Gulf of Mexico and the Atlantic Ocean. Departure from the climatological flow is apparent in the divergence anomaly east of the Sierra Madre and Rocky Mountain chains; i.e. the convergence has shifted eastward. The precipitation anomalies during these events are associated with anomalous large-scale convergence of moisture that stretches along the entire Mississippi and Ohio Valleys towards the Great Lakes.

To better understand the inherent characteristics of the moisture fluxes and convergence that caused the floods we decomposed the total daily moisture transport anomaly ($Q'_{\mathbf{v}}$) into components:

$$Q'_{\mathbf{v}} = -\frac{1}{g}\left[\int_{p_s}^{600} \bar{\mathbf{v}} q' \, dp + \int_{p_s}^{600} \mathbf{v}' \bar{q} \, dp + \left(\int_{p_s}^{600} \mathbf{v}' q' \, dp - \int_{p_s}^{600} \overline{\mathbf{v}' q'} \, dp \right)\right] \quad (1)$$

where p_s is the surface pressure in mb, g is the acceleration due to gravity, $\mathbf{v}$ is the vector velocity, q is the specific humidity, the overbars indicate the 1961-1990 climatological daily mean for the particular calendar date, and the primes denote the daily deviation from the climatological daily mean. Thus, the first right hand side term is the component due to the climatological circulation advecting the anomalous moisture field (hereafter referred to as the humidity anomaly term) and the second term is the component due to the circulation anomaly advecting the climatological moisture field (hereafter the circulation anomaly term), respectively. The last two terms are the nonlinear cross terms contributing to the total convergence. They depict the anomalous advection due to the covariance between anomalies of both the circulation components and the specific humidity with respect to the climatological eddy covariance term (hereafter the eddy covariance anomaly term).

The results of this moisture transport decomposition and the associated convergence are shown in Figs. 17.4 and 17.5 and correspond to the composite of the 20 historical flood events (Fig. 17.4), and the April 2011 event (Fig. 17.5), respectively. The terms are averaged over the 9-days leading to the flood and the flood dates itself (i.e., a 10 day

average). The humidity anomaly term contribution is shown in the top panel, the circulation anomaly term in the middle, and the eddy covariance anomaly term contribution in the bottom panel. The total anomalous moisture transport and the total anomaly in transport convergence (shown in the left panels of Fig 17.3) are clearly dominated by the circulation anomaly term (Figs. 17.4b and 17.5b). In combination, these figures indicate that the averaged 10-day rainfall anomaly, and subsequently the floods themselves, were caused by a unique, persistent and strong atmospheric anticyclonic circulation anomaly, east of the U.S. central Atlantic coast, which shifts the moisture transport from its typical path along the eastern foothills of the Rockies into the Mississippi/Ohio Valley region. This eastward shift weakens the climatological moisture convergence in the western high plains and generates anomalous moisture convergence in the east, against the Appalachians. The humidity anomaly term (Figs. 17.4a and 17.5a) is a direct response to the change in circulation, which dries the atmosphere over the western plains and increases the moisture to the east. In general, the high-frequency transient eddies respond by acting to damp the anomaly due to the change in circulation (Figs. 17.4c and 17.5c).

Individual daily components, as in equation (1), for selected days of the April 2011 event are shown in Fig. 17.6. These show that in the days immediately prior to the flood the eddy covariance term is stronger than when averaged over the prior 10 days (Fig. 17.5). The connection to the oceans is due primarily to the circulation anomaly alone, which is also responsible for channeling the oceanic moisture (supplied by full field circulation and moisture divergence, see Fig. 17.2 left side) into a corridor that leads to the flooded basins.

To further analyze the dynamical causes of the floods, we show in Fig. 17.7 maps of 700 mb geopotential height anomalies in contours and vertical pressure velocity in colors (blue/purple is upward motion) during days -9, -5, -3, and -1 prior to the flood date for the composite of 20 historical events (left) and the April 2011 event (right). The dipole pattern of a significant positive geopotential high anomaly to the east of the flooded basins together with a weaker low anomaly to the west, is established on day -9 and persists throughout the days leading to the flood event. Consistent with quasi-geostrophic dynamics, there is northward and upward motion over the flooded basins. Flow around the high can be seen in the vectors of moisture flux on Fig. 17.3. Both the historical and April 2011 events show an intensification of the dipole between day -9 and day 0, and transition from southwesterly to a more direct southerly flow into the basin in the days leading up to the flood.

To emphasize the persistent moisture transport and precipitation that are responsible for these flood events, we plot time series of anomalous precipitation, temperature, vertical velocity, integrated water vapor, northward water vapor transport (Q_v) and lower tropospheric and near surface moisture convergence, all averaged over the flood basins (Fig. 17.8 a-d) between 100°W and 90°W, and 28°N to 42°N (Fig. 17.8 e and g) from days -10 to day 2, where day 0 corresponds to when the 10-year return threshold is exceeded. For both historical floods (grey lines, shading) and the April 2011 flood (black line). There is a positive trend in all variables until one day before the flood, or on the day of the flood, and a steep drop thereafter. Note that during the days leading to the April 2011 flood, there is a break in the upward trend between 7 and 6 days before the flood

date. On those days the anomalous flow from the south at 700 mb shifts to the west briefly (not shown) taking the flood basin out of the area of high moisture transport convergence. Breaks in precipitation in the days leading up to a flood are standard and are associated with the movement of high-frequency synoptic events that move through the more permanent quasi-stationary dipole that is associated with the floods. The break does not appear to be significant when averaging 20 events due to the stochastic nature of the underlying synoptic variability. The moisture transport, convergence and precipitation reach their maximum values a couple days prior to the flood. The delay of stream flow relative to precipitation may be due to the travel time of runoff water through the basin towards the gauging stations. Precipitation (Fig. 17.8a) is not exceptionally high on any one day, making the duration important for flood generation through surface soil saturation and subsequent production of overland flow. The basin averages of near surface temperature, 700 mb vertical velocity, and integrated water vapor (Fig. 17.8 b-d) show low-level warm, moist (and presumably buoyant) air in the region from days -9 to -7 and again from days -5 to the flood date. The narrow strip between 100°W and 90°W from the coast to the northern edge of the basin (Fig. 17.8 e and g) was selected to show the Great Plains meridional moisture flux (GP MMF). Fig. 17.8e shows just the meridional part of the surface - 600 mb vertically integrated moisture transport (Q_v) and indicates the transport of warm moist air from the tropics. Also shown, in Fig. 17.8f, is the 700 mb geopotential height difference between the centers of the anomalous high pressure cell to the east and the low pressure cell to the west of the Ohio Basin (designated hereafter as the 'dipole index'), defined based on the locations of these centers on day -1 (see Fig. 17.7). This difference is proportional to the geostrophic flow

in between. The persistently positive anomalous pressure gradient before the flood drives anomalous southerly flow into the region as seen in both the historical and the April 2011 floods. Panel 8g shows 600 mb – surface convergence consistent with lifting of this air within the narrow northward flowing moisture band from days -9 to -1.

The eastern North American double maximum of poleward moisture transport identified by Knippertz and Wernli (2010) with channeling of moist flow between the Rocky and the Appalachian Mountain chains (the Great Plains meridional moisture flux, GP MMF), and a Gulf Stream flow (GS MMF) to the east is visible in Figs. 17.9 and 17.10, which show a Hovmöller, longitude-time plot of vertically integrated surface - 600 mb meridional wind (contours) and moisture transport (colors) anomalies along the latitude of 25°N (Fig. 17.9) and 35°N (Fig. 17.10) for the composite historical flood events (top) and the April 2011 event (bottom). The cross section of the surface elevation is plotted for reference but note that for the meridional winds and moisture transport the vertical axis marks time in days with reference to the flood date. Maximum meridional transport at 25°N is located at days -4 to 0 for the historical composite, and there is a maximum at day 0 for the April 2011 event. Alignment of contours and colors shows the dominant control of the circulation anomaly on the total moisture transport anomaly. Moving from 25°N to 35°N there is an intensification of the wind and a corresponding intensification of the moisture transport. Maximum meridional transport at 35°N is located at day -2 for the historical composite GP MMF and day -1 for the GS MMF, and there is a double maximum at days -8 and days -1 through day 0 of the GP MMF for the April 2011 event with the GS MMF maximum a day after. Note that the anomalous moisture transport occurs at the eastern edge of the climatological springtime moisture transport path (Fig.

17.1) indicating a shift eastward due to the anomalous circulation patterns shown in Fig. 17.3. Therefore, the strong moisture transport into the Ohio basin is anomalous and linked with the flood events exceeding the 10 year return period.

Many of the definitions of the atmospheric river phenomenon set by Ralph and Dettinger (2011) with reference to the Pacific-North America sector are met in these cases of Ohio River Basin floods. In particular, integrated water vapor greater than 2 cm, wind speeds greater than 12.5 ms^{-1}, and a narrow width of the moisture flow. The cross-section in Figure 17.10 shows that at 35°N the moisture flow is approximately 400-500 km across. The Ohio River moisture transport does not meet the length criterion of Ralph and Dettinger in a strict sense, but the distance from the Ohio River region to the Gulf of Mexico, Caribbean Sea, and the tropical Atlantic is naturally much smaller than the Pacific Ocean, which can support long plumes of moisture from the central Pacific into the west coast of North America.

To explore the longer timescales associated with these flood events, the daily dipole index was computed and ranked for the months of January through May and from 1901 to 2008. From this dataset, the ranked values of the dipole index were composited day-by-day for days -59 to +5 for the historical 20 events (Fig. 17.11). The composited values (shown in percent) indicate that during the 65-day period, the dipole values fluctuated, on average, between the 40^{th} and 80^{th} percentile of all historical daily values. Fig. 17.11 shows a pattern that builds up to successively higher percentile value up to the flood event. The figure suggests that there are two time scales in the buildup: a short one that roughly corresponds to the synoptic time scale of ~10 days and a longer one with a ~ one month time scale. The phase locking of these "cycles" with respect to the peak flood date

indicates the role of baroclinic waves as well as possibly longer, subseasonal oscillations that may play a role in the lead-up to these extreme floods. Kushnir and Wallace (1989) noted the existence of a waveguide for subseasonal oscillations, extending from the eastern North Pacific, into North America, towards the western North Atlantic, with a structure similar to the patterns shown in Fig. 17.7. Possible origins for such variations are the tropical (Madden Julian Oscillation - MJO) intra-seasonal oscillation, which is known to impact North America (Higgins and Mo 1997), and 20–30 day mid-latitude oscillations associated with the Rockies (mountain torques) that exhibit circulation structures similar to those of the dipolar pattern in Fig. 17.7 (Lott et al. 2004). Evidence that the MJO impacts weather types associated with Midwest flooding events is reported in Robertson et al. (2012).

15.2 Dipole Index and Extreme Precipitation Relation

Using GPCC station precipitation and NCEP CPC regional daily gridded precipitation averaged over the Ohio basin, daily events over the 99th percentile were about ten times more likely in spring when the dipole index was greater than 150 mb. Figure 17.12 shows the scatter and the percentile regression quantiles (Koenker and Bassett Jr. 1978) from 0.1 to 0.9 of the dipole index in mb and the precipitation in mm day^{-1} averaged over the Ohio basin for the 1901-2008 GPCC station data (left) and 1948-2004 NCEP CPC regional daily gridded data (right). Both the scatter data in dots and the regression quantile lines show an overall trend of increasing values of precipitation with increasing values of dipole index. The NCEP data (right) is a shorter time series, but may have a higher data reliability than the GPCC (left). These trends are explored further by the

quantile regression in Figs. 17.13 (precipitation greater than 50th, 75th, and 90th percentiles), and 17.14 (precipitation greater than 95th, 98th, and 99th percentiles) as expressed as the conditional probability on the y-axis and dipole index in mb on the x-axis. In Figs. 17.13 and 17.14 the green line is a linear model with the logistic regression, and the red line is a nonparametric logistic regression using B-Splines and automatic knot selection. Box plots of precipitation and dipole index along with the density are located on the axis of Figs. 17.13 and 17.14. As in Fig.17.12 the GPCC data is on the left and the NCEP data is on the right. For precipitation greater than the 50th percentile (top Fig. 17.13), areas of significance are indicated when the conditional probability is far from the expected value of 0.5. Similarly in the 75th percentile (middle Fig. 17.13) areas far from 0.25, 90th percentile (bottom Fig. 17.13) far from 0.1, 95th percentile (top Fig. 17.14) far from 0.05, 98th percentile (middle Fig. 17.14) far from 0.02, and 99th percentile (bottom Fig. 17.14) far from 0.01. For the 99th percentile (Fig. 17.14 bottom) the probability is almost 0.1 not 0.01 as we would expect for the dipole index greater than 150 mb, i.e. 10 times as likely as by chance. The GPCC (left) and NCEP data (right) show the largest differences in Fig. 17.14 as the percentile goes up. The longer data set (GPCC on left) has a recurve of B-Spline regression indicating a maximum probability at about 200 mb for the dipole index, while the shorter data set (NCEP on right) has no upper limit.
The daily dipole index is not normalized so the stronger high pressure dominates. Although low pressure is a necessary component of flooding providing the upward motion, it is the blocking high that is the signature on all the 10-year return period events in the Ohio basin. A study compositing cool season events in Newfoundland (Milrad et. al. 2009), winter southeast U.S. (Katz, 2003) and summer southeast U.S. (Li et. al., 2011)

found a anomalous blocking high to the east of the basin as a signature of precipitation events. Lu et. al (2013) also found a blocked stationary wave pattern with a hemispheric empirical orthogonal function (EOF) featuring moisture transport from the Atlantic for an extreme flood in France.

15.3 The general role of circulation anomalies

The Ohio basin floods are clearly caused by atmospheric circulation anomalies with a largely passive advection of water vapor, which condenses out as rainfall over the basin. However the Ohio basin floods are just one case study in a particular environment strongly influenced by the Bermuda High and the topography of the Rockies and Appalachians. To assess whether the local characteristics of the dynamic circulation control is more general, we turn to a very different meteorological and climatological environment, that of Great Britain to the immediate east of the Atlantic Ocean and with prevailing southwesterly winds. Lavers et al. (2011) identified winter atmospheric rivers originating in the subtropical Atlantic and flowing to Great Britain as responsible for severe flood events in a number of Britain's river basins. One basin they examined was that of the River Eden in northwest England where they identified ten flood dates at Temple Sowerby. Eight of these flood events occur prior to 2008 (7 Jan 2005, 3 Jan 1982, 31 Jan 1995, 28 Dec 2007, 23 Feb 1991, 3 Feb 2004, 5 Dec 2006, and 21 Dec 1985). For these events we conducted the same form of moisture transport decomposition as done for the Ohio basin, taking the day -5 to day 0 average (the shorter averaging interval corresponding to the shorter duration of these events). The results in Fig. 17.15 are shown in a similar manner to Fig. 17.4. Over Great Britain, as in the cases

of the Ohio basin, the total moisture transport anomaly and convergence are dominated by the circulation anomaly term (Fig. 17.15b), which depicts a pattern that shifts the moisture flux convergence from its normal location in the mid-Atlantic in the box 35°-50°N and 60°-30°W (Fig. 17.1) northeastward, towards the British Isles. The consequential moisture deficit created in the climatological convergence region and excess over the British Isles explain the moisture anomaly component of the flux and its convergence (Fig. 17.15a). Here too the eddy covariance term acts to damp the circulation anomalies. This indicates that the floods, as in the Ohio basin, were caused by persistent and strong circulation anomalies. Figure 17.16 shows the individual daily terms, as in equation (1), for days -2 and -3 for the Great Britain winter flood event of 3 February 2004. As in the Ohio River Basin, the eddy covariance anomaly term is stronger near the basin on the daily rather than the time-averaged time scale but the connection to the subtropics is due primarily to the circulation anomaly. The Day -3 (Fig. 17.16, panel b) circulation anomaly shows a striking direct link from the Caribbean to Great Britain.

The concentration of the 900 mb specific humidity (as shown in Fig. 17.3 of Lavers et al. 2011) into filament like, atmospheric river structures seems to apply in the Ohio River case as well. To show this we follow Lavers et al. (2011), and present the 900 mb specific humidity fields one day prior to four of the 20 historical Ohio basin flood dates to illustrate the GS MMF (Fig. 17.17) and two days prior to illustrate the GP MMF (Fig. 17.18). These maps indeed show some similar filament like structures for both GS and GP MMF. 16 May 1943 is a representative example of a few dates in which both GS and GP MMF are present at day -2 with a small separation between the two. Surface – 600

mb integrated moisture convergence maps of these days (not shown) mirror the locations of the 900 mb specific humidity maximums.

This page intentionally left blank

Chapter 16

Summary and Conclusions

We have identified important spatial and temporal features of the atmospheric circulation that lead to the extreme spring floods in the Ohio River Basin. This was done by compositing meteorological fields for 20 historical events and, separately the more recent April 2011 event. Common to all the floods is the low level moisture divergence (source) in the tropical Atlantic, which is a climatological feature. Scattered anomalous sources are located in the northern Gulf of Mexico and are due to the anomalous steering of the climatological moisture supply. The anomalous low-level convergence (sink) above the flood basins is associated with a reproducible, persistent, anomalous circulation feature accompanied by strong upward vertical motion over the Basin. The anomalous circulation is driving a strong southerly low-level flow of warm and moist air into the Ohio Basin, channeled from the Atlantic and against the Appalachian mountain chains between an anomalous quasi-stationary high- and low-pressure cells to the east and west, respectively. This feature begins to develop as much as nine days before the flood date, with stronger features established five days before the event. This anomalous pattern (Fig. 17.3, right panels) is an eastward shift of low-level convergence and moisture flux compared to the climatological presented in Fig 17.1. The “Maya Express” in Dirmeyer

and Kinter (2009) is, in contrast, more akin to the intensification of the climatological pattern of moisture transport to a region west of the Ohio basin and associated with a strengthening, or westward shift of, the Bermuda high. Instead, during extreme Ohio basin floods, the anomalous high in the Atlantic is north of the Bermuda high and is paired in a dipole fashion with a low to the west of the Basin on the western flank of the climatological high. The 20 flooding events show remarkable similarity in circulation pattern and the fact that the composite of the 20 events is still coherent after averaging shows the consistency of this pattern. The 2011 flood was also found to be typical of previous floods in the basin. This study is a sound confirmation of the Hirschboeck (1991) hypothesis that extreme floods in a given large basin and at a given season are due to typical circulation types and are not a collection of random, unrelated events. Decomposition of the anomalous total moisture transport associated with Ohio Basin floods into the transport of climatological moisture by the anomalous atmospheric wind circulation and the transport of anomalous specific humidity by the climatological circulation confirms that these floods are governed by the anomalous circulation. Similar decomposition of winter atmospheric rivers impacting the River Eden at Temple Sowerby in northwest England shows these also are caused by atmospheric circulation anomalies, though those are different than in the Ohio case. This makes it clear that for these very different meteorological, climatological and geographical environments, the strong moisture fluxes and the floods they cause, are fundamentally anomalous dynamical features of the atmospheric circulation (and not of the moisture field) and, may be due to quasi-geostrophic circulations intensifying gradients along a front, as described

by Hoskins and Bretherton (1972). This 'atmospheric channeling' of the moisture field could be referred to as an atmospheric river.

The persistent meteorological patterns associated with the Ohio floods, suggest the possibility that global-scale circulation driven by specific non-local dynamics, may contribute to them. For example, the long-lasting La Nina climate pattern, persisting into the spring of 2011, could have contributed to the basic large-scale ingredients of the event, with a strong cold trough over Canada and warm ridging over southwest US and Atlantic Ocean. The daily circulation types associated with Ohio Basin flooding events were found to be significantly more prevalent during La Niña events by *Robertson et al.* (2012). However the March to May NINO 3.4 index averaged for the years of the 20 historical events is only trivially negative (-0.02 °C) so the relationship with ENSO is weak in general.

Interestingly, Ohio River floods were more common in the early and mid-20th Century than in recent decades. Whether the 2011 event heralds a return of more frequent flooding, or whether this is merely a result of random variability, requires more research into how the frequency and intensity of the synoptic events responsible for the floods varies on decadal timescales and how they are potentially influenced by large-scale modes of low frequency climate variability.

This page intentionally left blank

Chapter 17

Figures and Tables

Figure 17.1: March-May 1961-1990 climatology of vertically integrated surface - 600 mb moisture flux in kg m^{-1} s^{-1} (strongest 20 percent of values shown as arrows) and moisture convergence in mm day^{-1} (colors) with basin outlines in hot pink.

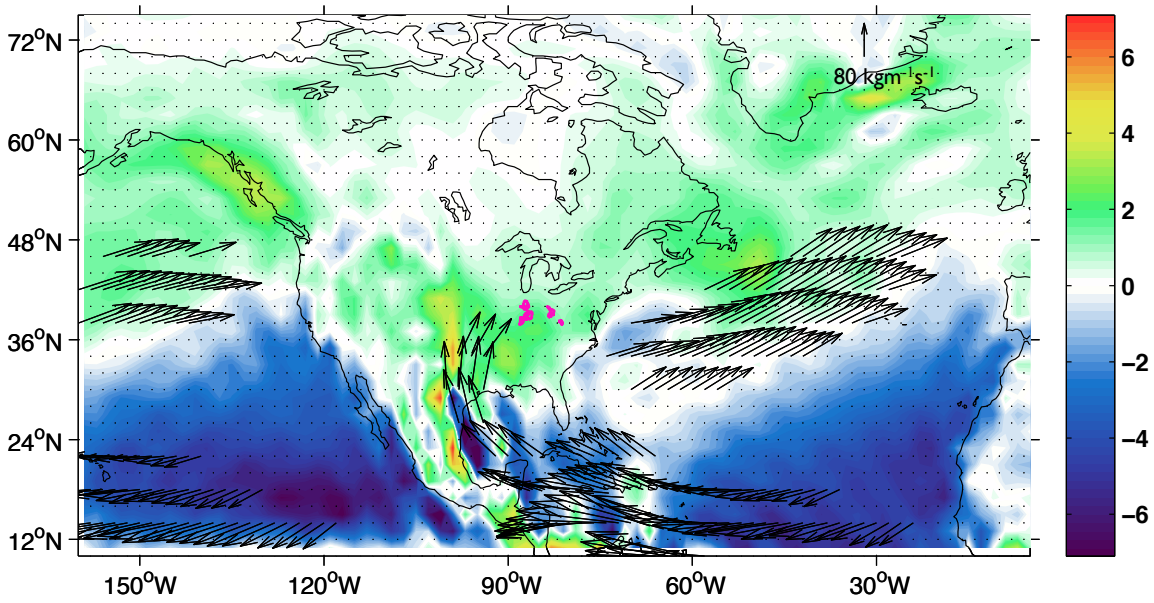


Figure 17.2: Precipitation anomaly in mm day^{-1} of a) National Climatic Data Center's Global Historical Climatology Network (GHCN) daily station data gridded over land into one-degree boxes average of 20 historical, 20th century floods in large river basins in the Ohio Valley (basin outlines in hot pink) and b) NCEP-NCAR Reanalysis average of 18-27 April 2011.

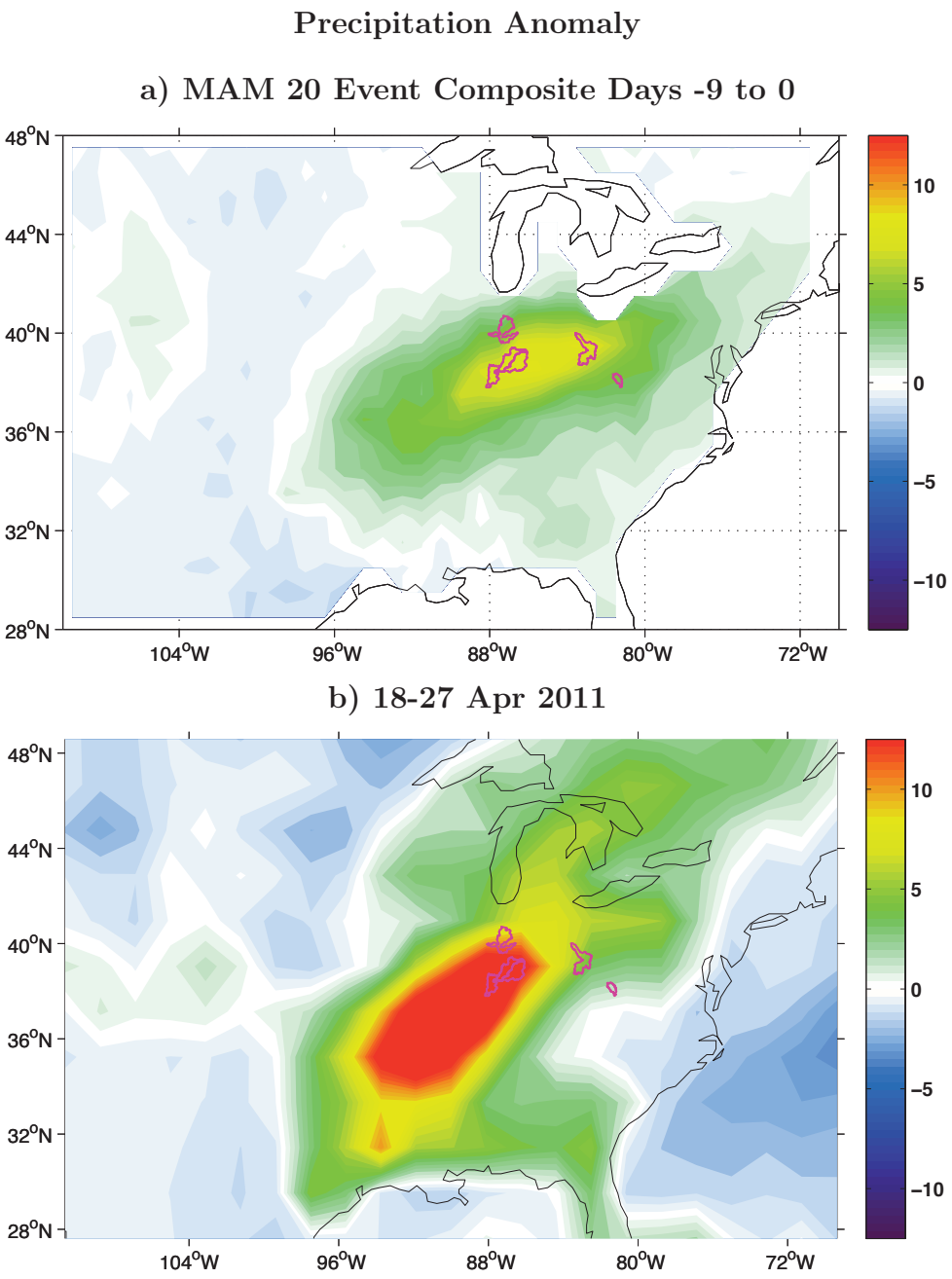


Figure 17.3: (a) Vertically integrated surface - 600 mb moisture flux in kg m^{-1} s^{-1} (strongest 20 percent of values shown as arrows) and moisture convergence in mm day^{-1} (colors) for composite 20 extreme flood events observed in large drainage basins (size > 10^3 km^2) within the Ohio Valley (basin outlines in hot pink) averaged over the nine days leading to the 10-year flood. (b) same but for 18-27 Apr 2011. Full field (left) and anomaly relative to a 1961-1990 climatology (right).

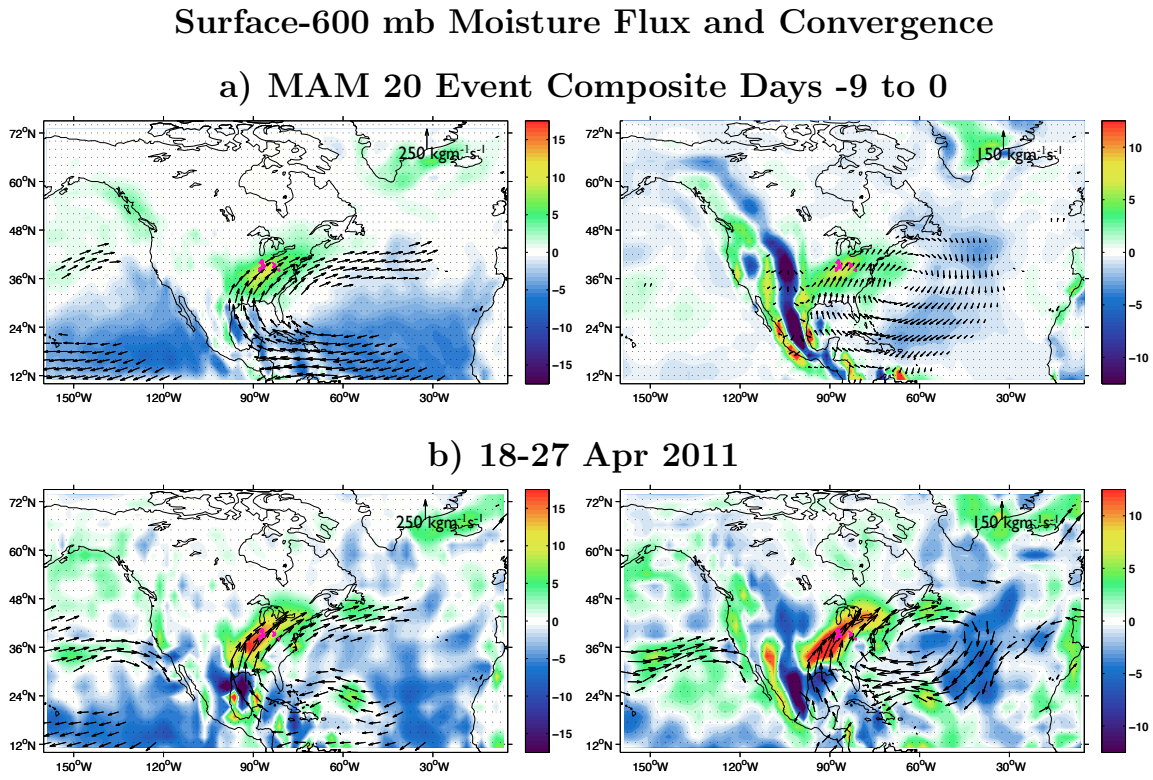


Figure 17.4: Decomposition of the anomalous moisture transport for the 20 historical composite averaged over the nine days leading to the flood (basin outlines in hot pink). Moisture flux in kg m^{-1} s^{-1} (strongest 20 percent of values shown as arrows) and moisture convergence in mm day^{-1} (colors). (a) Climatological circulation transporting the moisture anomaly (humidity anomaly term), (b) circulation anomaly transporting the climatological humidity (circulation anomaly term), and (c) circulation anomaly transporting the moisture anomaly (eddy covariance anomaly term).

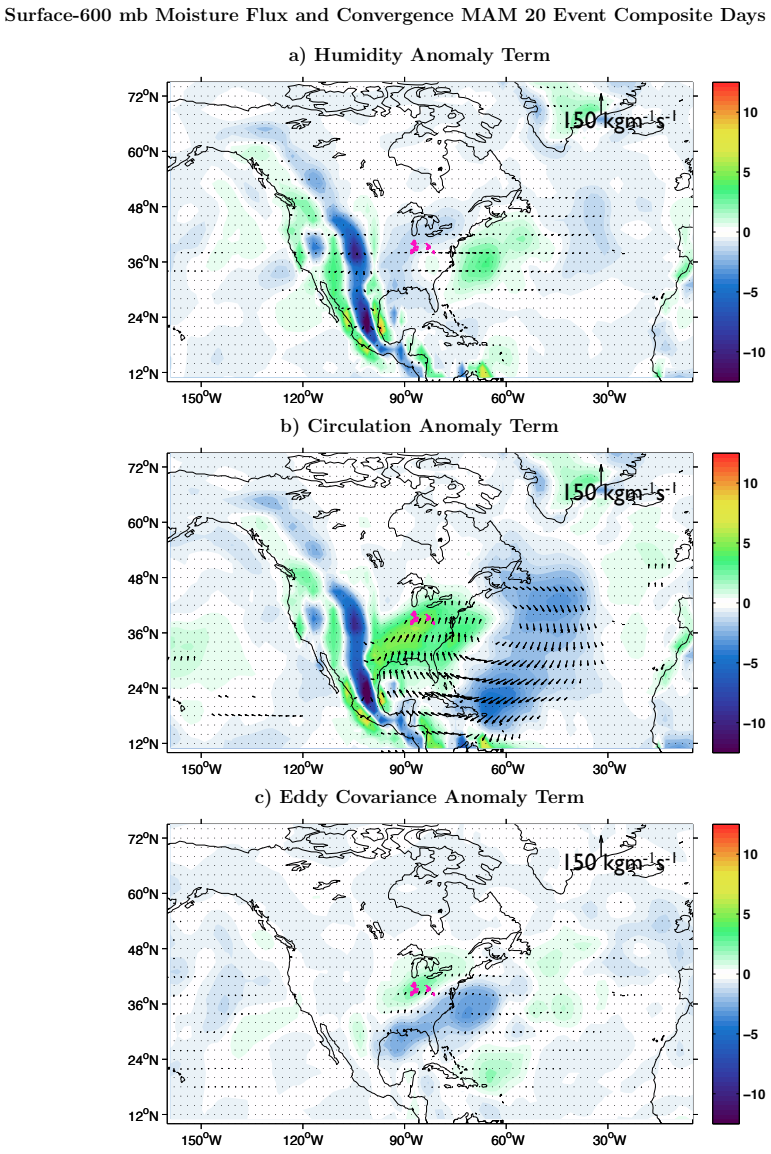


Figure 17.5: Same as in Fig. 17.4, but for the April 2011 event.

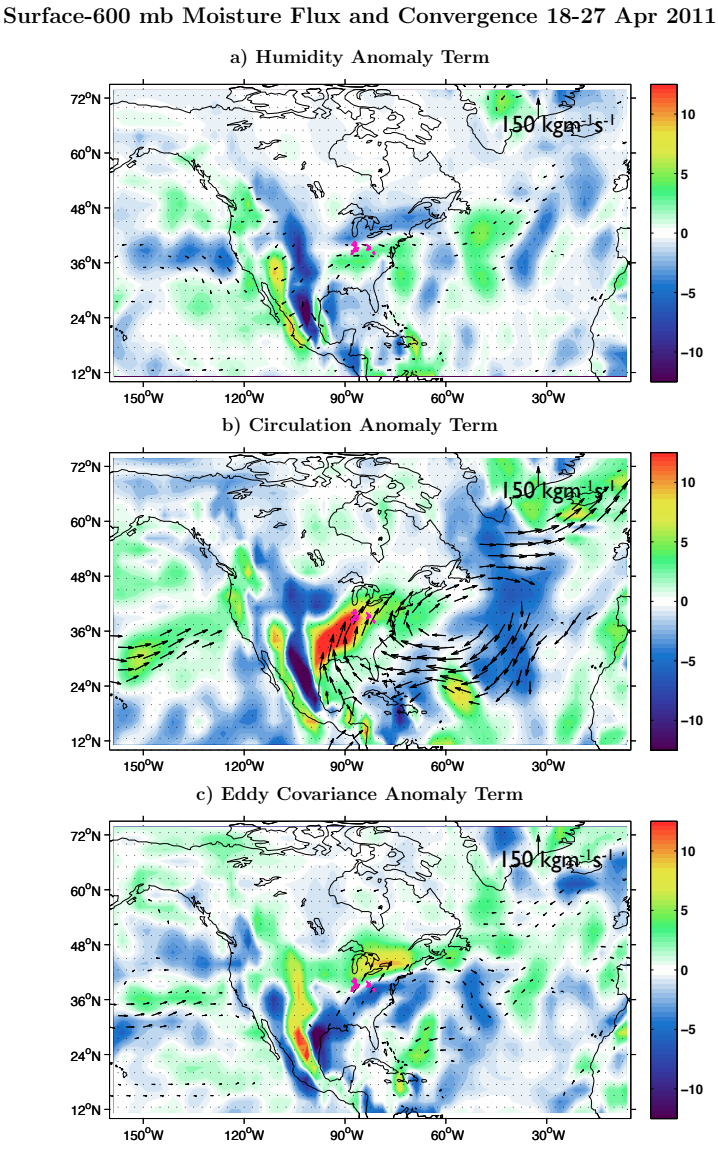


Figure 17.6: Daily decomposition of the anomalous moisture transport for the April 2011 event (basin outlines in hot pink). Moisture flux in kg m^{-1} s^{-1} (strongest 20 percent of values shown as arrows) and moisture convergence in mm day^{-1} (colors), (top) humidity anomaly term, (middle) circulation anomaly term, and (bottom) eddy covariance anomaly term. (a) Day 0: 27 Apr 2011 and (b) Day -2: 25 Apr 2011.

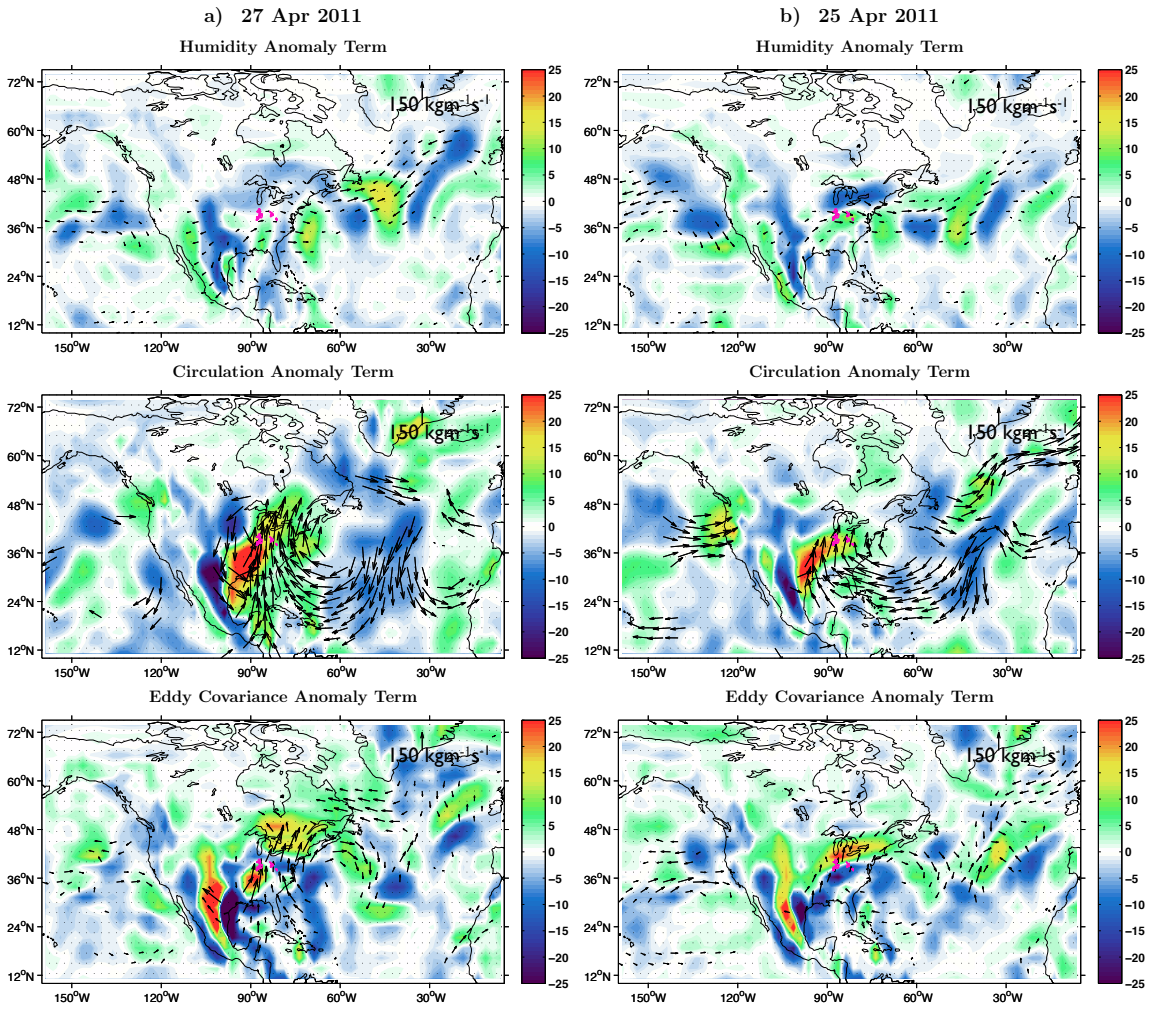


Figure 17.7: Average of 20 historical, 20th century floods in large river basins in the Ohio Valley (basin outlines in hot pink) in left column and days 18, 22, 24, and 26 of April 2011 in right column. Shown are 700 mb geopotential height anomalies (in m) in contours and vertical pressure velocity (in mb day^{-1}) colors (blue/purple is upward motion). Positive contours indicate anomalous positive heights (high pressure) and negative contours indicate anomalous negative heights (low pressure).

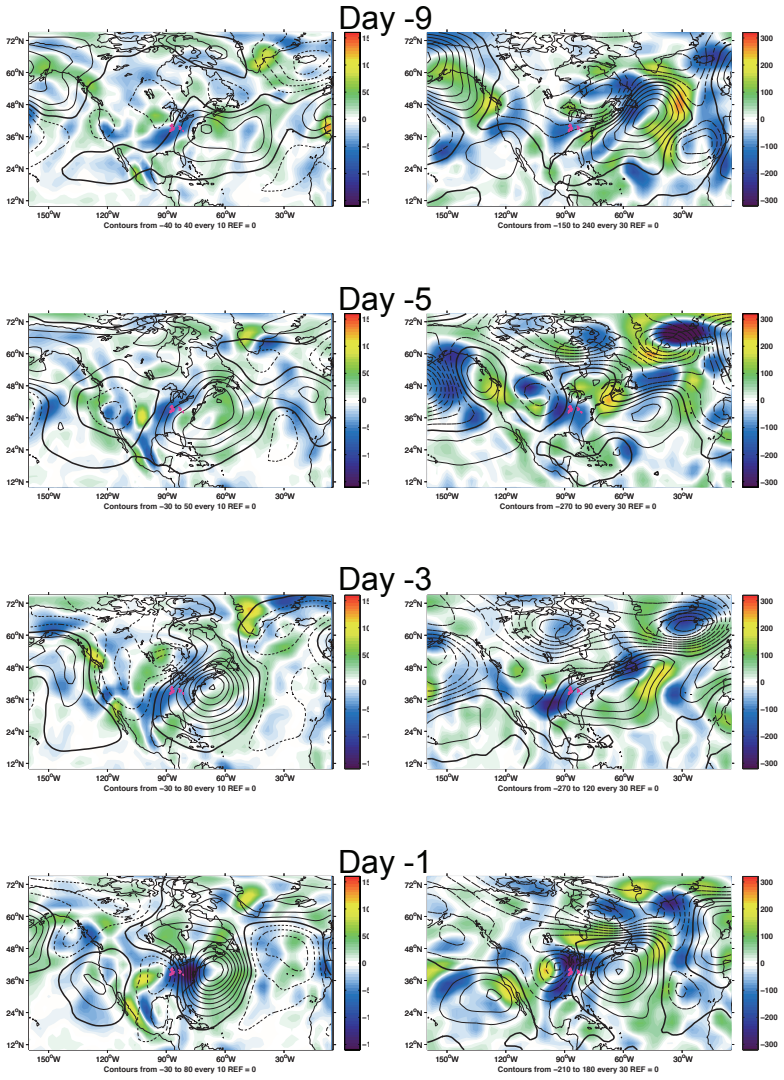


Figure 17.8: Anomalous time series of variables averaged over the Ohio River basin (a-d), between 100°W and 90°W and 28°N to 42°N (e and g), and height difference (f) from days -10 to day 2 (flood is day 0). Shown is the 20-event composite mean (solid grey) median (dashed grey) and 25th to 75th percentile spread (shaded grey) as well as the time series of the April 2011 event (black). (a) is the precipitation, (b) near surface air temperature, (c) 700 mb vertical pressure velocity, (d) integrated water vapor, (e) surface - 600 mb vertically integrated meridional moisture flux, (f) difference between 700 mb high and low height centers, and (g) surface - 600 mb vertically integrated moisture convergence.

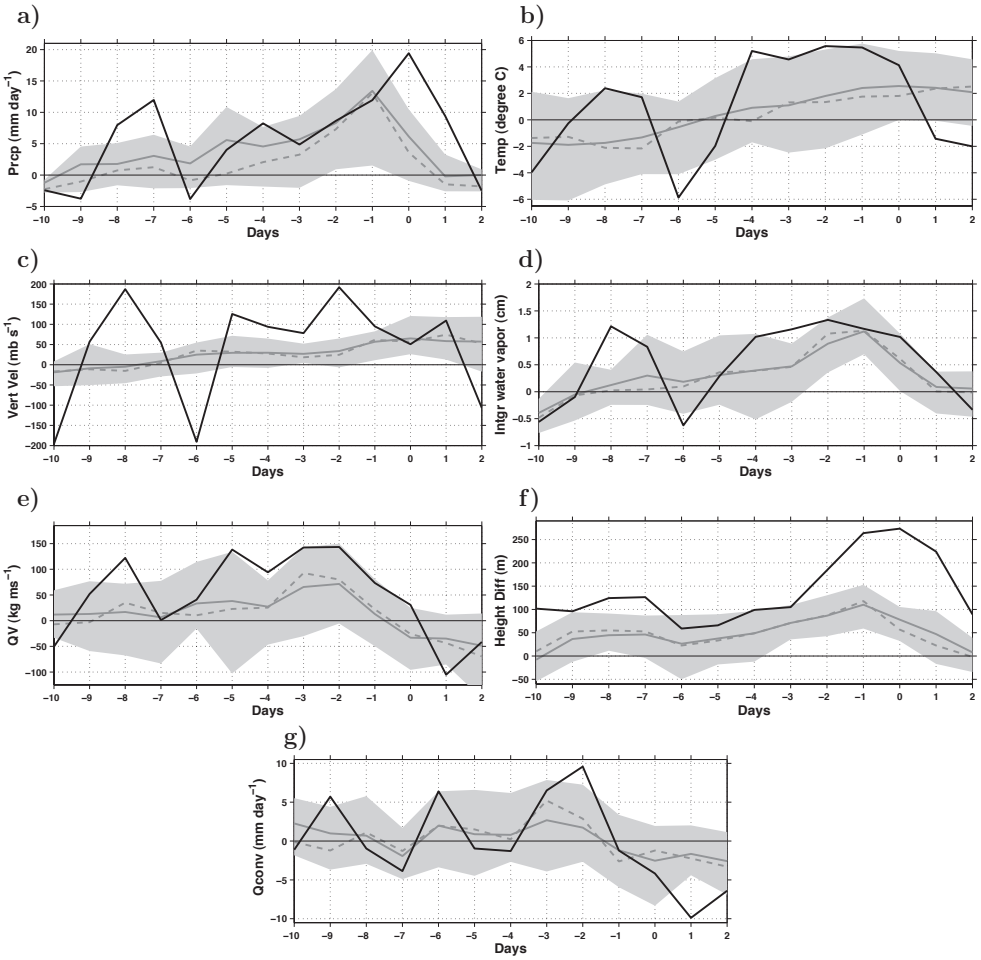


Figure 17.9: (a) Composite of vertically integrated anomalous surface - 600 mb meridional winds (contours mm day^{-1}) and moisture transport (colors in kg m^{-1} s^{-1}) along the latitude of 25°N for 20 extreme flood events observed in large drainage basins (size > 10^3 km^2) within the Ohio basin. They are plotted in days vs. longitude where day 0 is the first day of the flood (the first day on which the stream flow reaches the 10-year threshold). The cross section of the surface elevation is plotted for reference but note that for the contoured moisture transport the vertical axis marks the time with reference to the flood date. (b) Same as above but for the 27th of April 2011 flood event. Longitude range of Ohio sub-basins marked in red on x-axis.

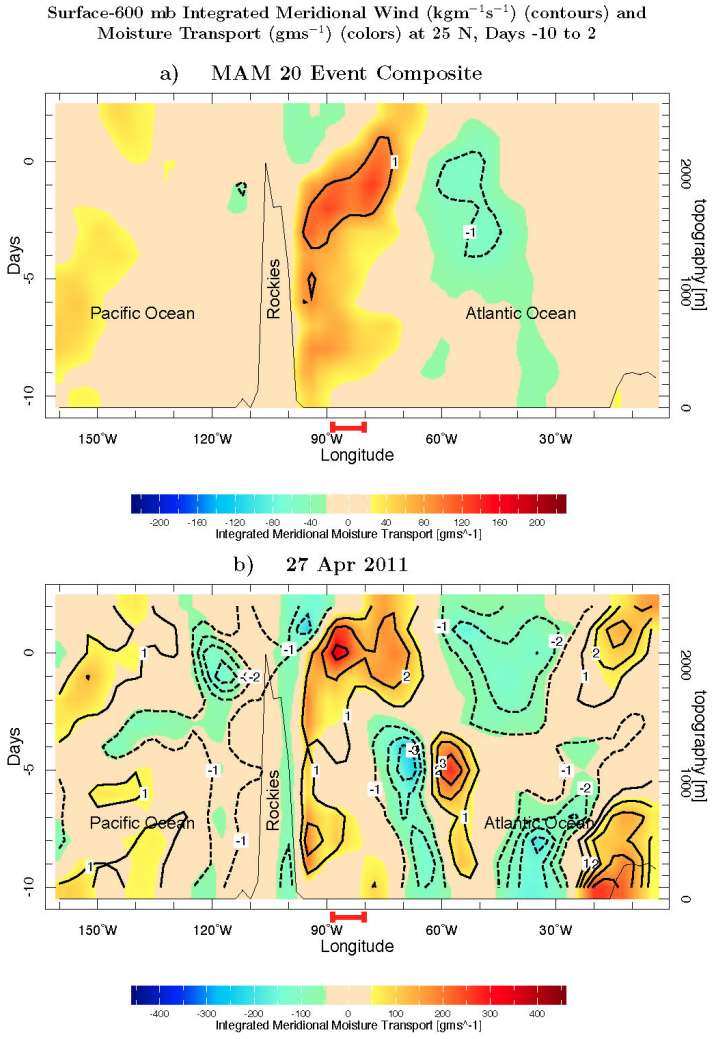


Figure 17.10: Same as for Fig. 17.9, but along the latitude of 35°N.

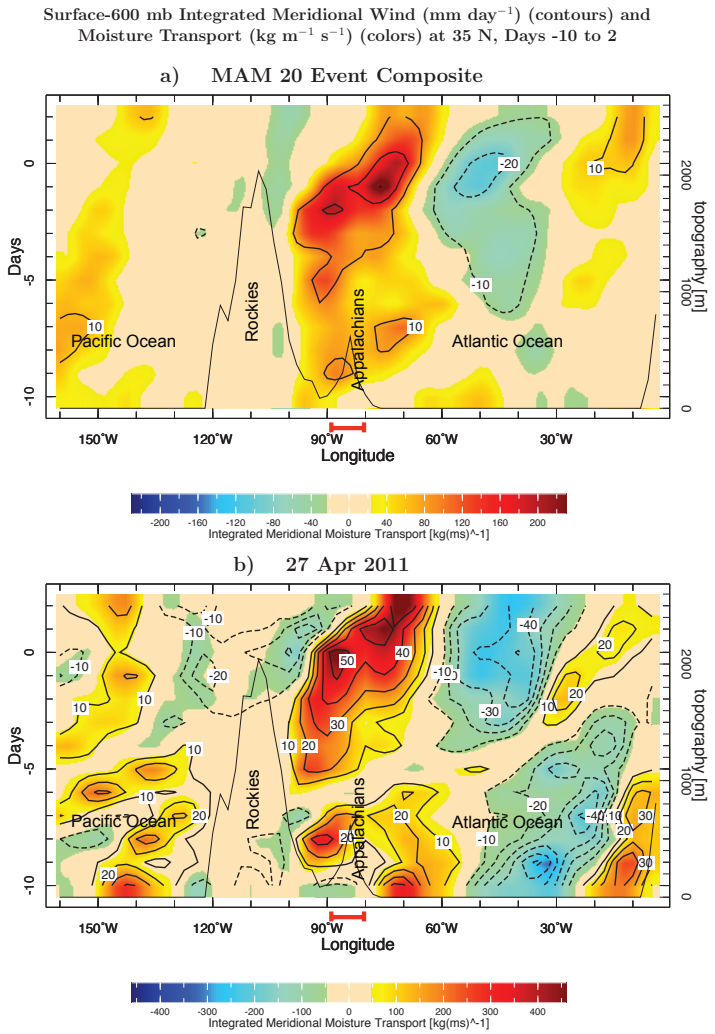


Figure 17.11: 20-event composite mean (solid), median (dashed), and 25th to 75th percentile spread (shaded grey) of days (-59 to 5, where 0 is the flood date) vs. percentile relative to a Jan-May 1901-2008 climatology of the dipole index. Dipole index computed by taking the largest geopotential height of the 700 mb anomalous high pressure system in the Atlantic Ocean minus the smallest geopotential height of the anomalous low pressure system west of Ohio basin at their respective locations on day -1.

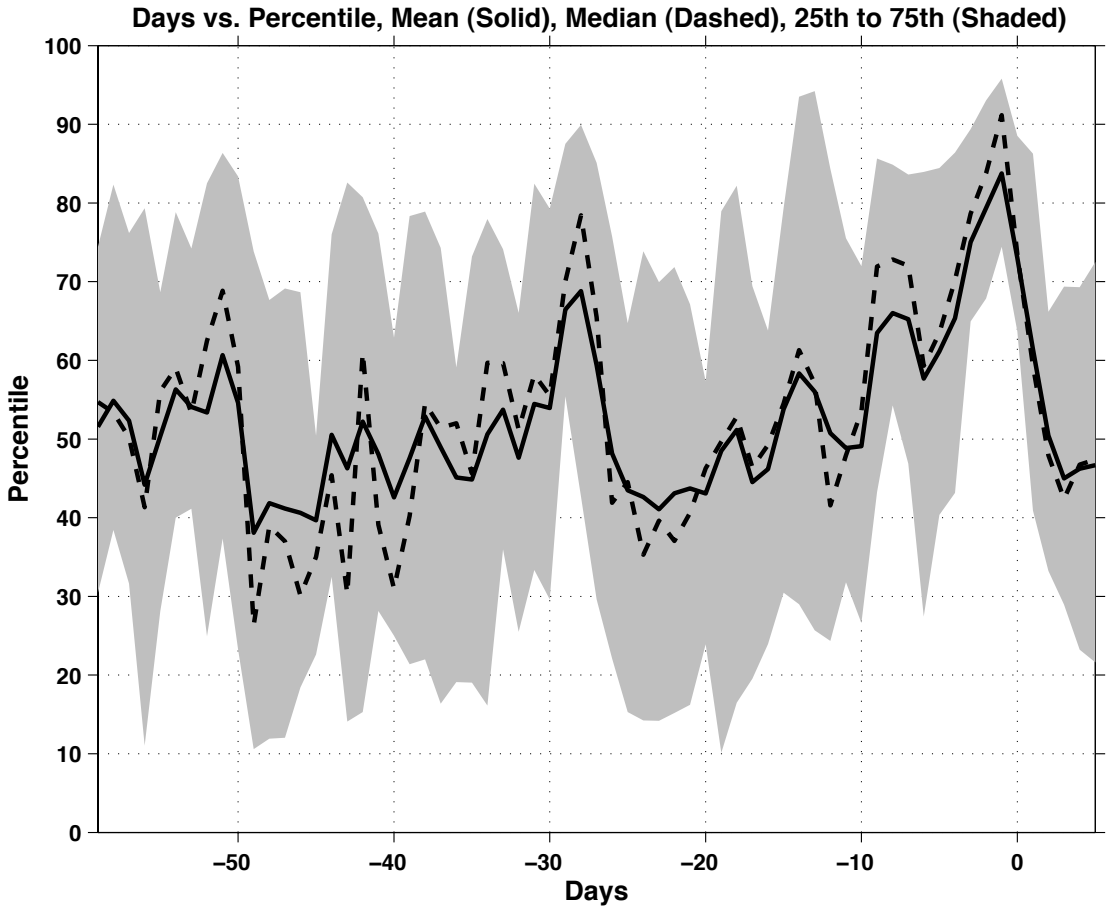


Figure 17.12: Scatter and the percentile regression quantiles (Koenker and Bassett Jr. 1978) from 0.1 to 0.9 of the dipole index in mb (as discussed in chapter 15) and the precipitation in mm day^{-1} averaged over the Ohio basin for the 1901-2008 GPCC station data (left) and 1948-2004 NCEP CPC regional daily gridded data (right). 0.1 solid black line, 0.2 dashed red, 0.3 dotted green, 0.4 dashed dark blue, 0.5 dashed cyan, 0.6 solid magenta, 0.7 dashed black, 0.8 dotted red, 0.9 dash-dot green.

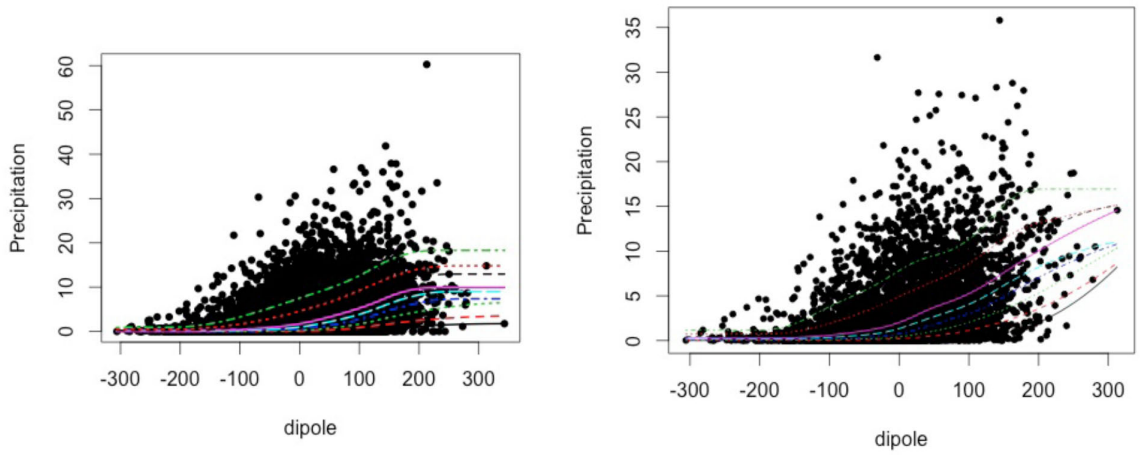


Figure 17.13: Quantile regression, as expressed as the conditional probability on the y-axis and dipole index in mb on the x-axis. Precipitation greater than 50th percentile (top), 75th percentile (middle) and 90th percentile (bottom). Green line is a linear model with the logistic regression, and the red line is a nonparametric logistic regression using B-Splines and automatic knot selection. Box plots of precipitation and dipole index along with the density are located on the axis. As in A.1 the GPCC data is on the left and the NCEP data is on the right.

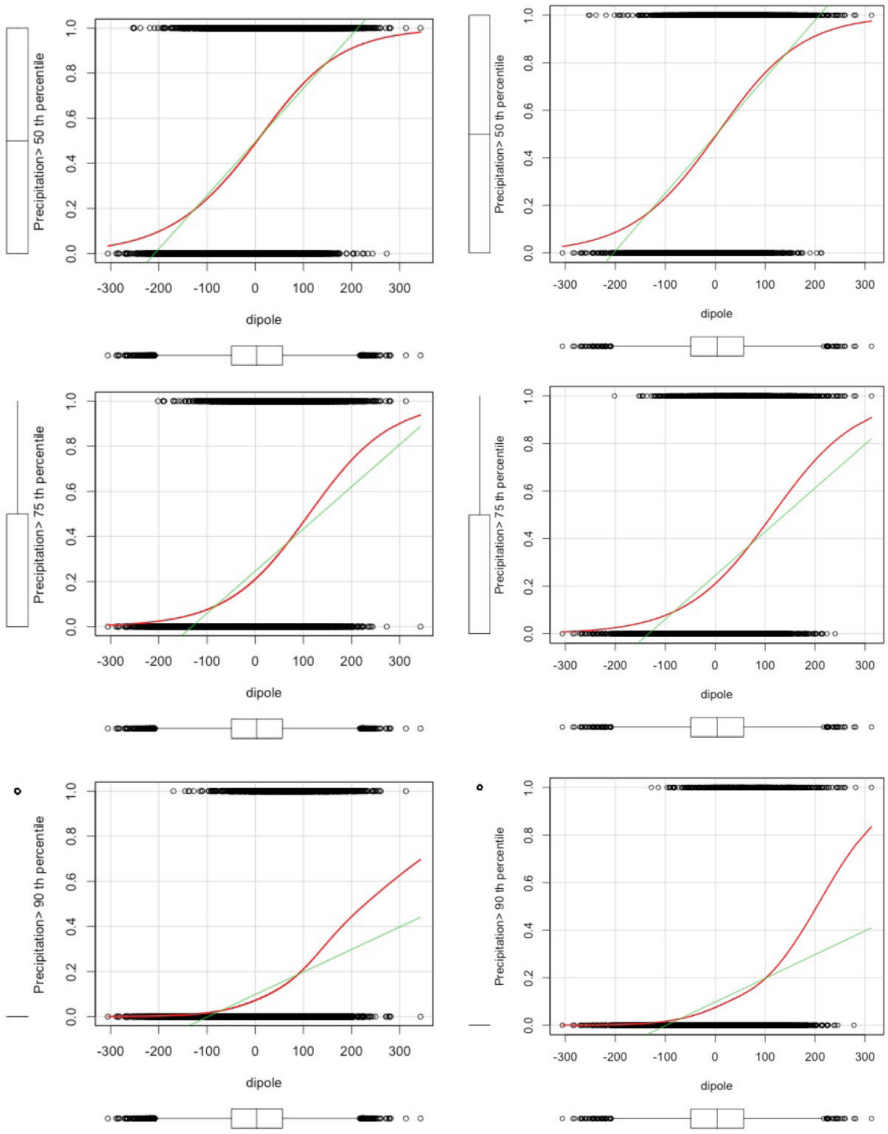


Figure 17.14: Same as 17.13, except precipitation greater than 95th percentile (top), 98th percentile (middle) and 99th percentile (bottom).

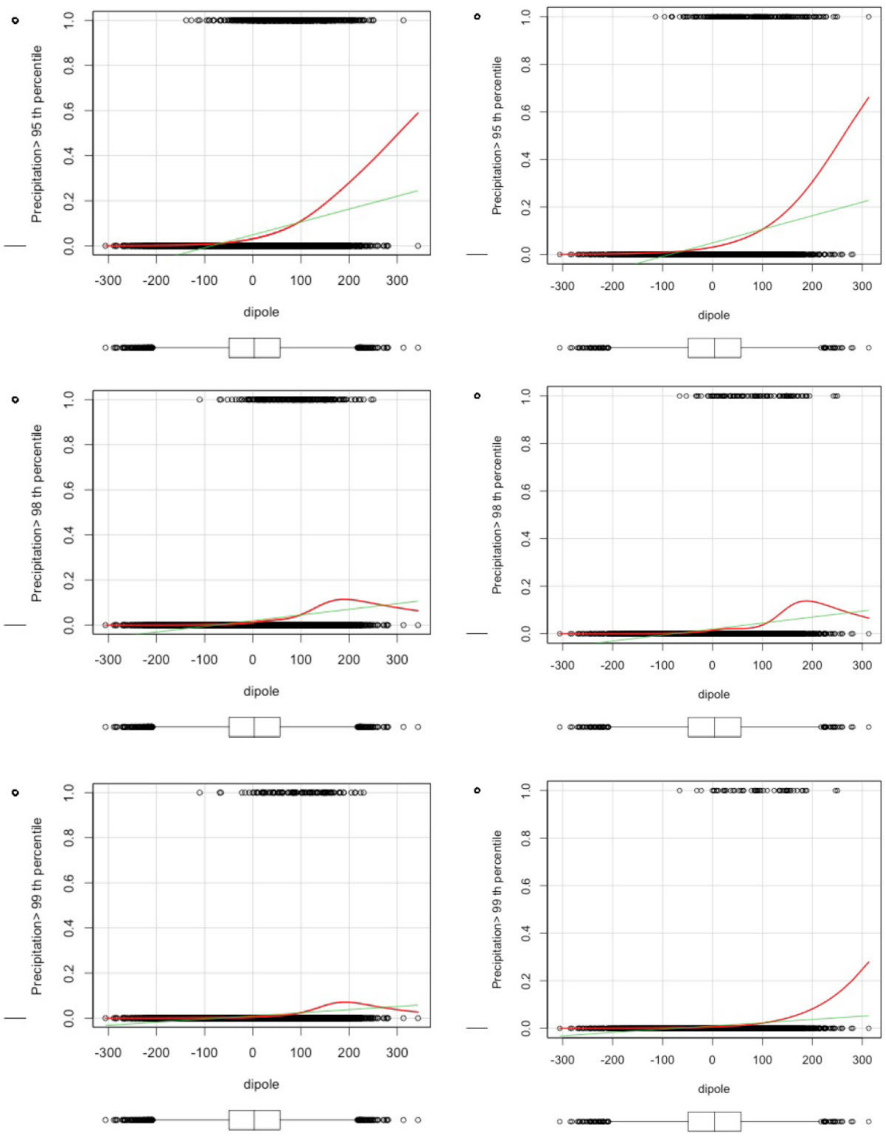


Figure 17.15: Decomposition of the anomalous moisture transport for eight flooding events floods on the River Eden at Temple Sowerby composite averaged over the five days leading to the flood. Moisture flux in kg m^{-1} s^{-1} (strongest 20 percent of values shown as arrows) and moisture convergence in mm day^{-1} (colors). (a) humidity anomaly term, (b) circulation anomaly term, and (c) eddy covariance anomaly term.

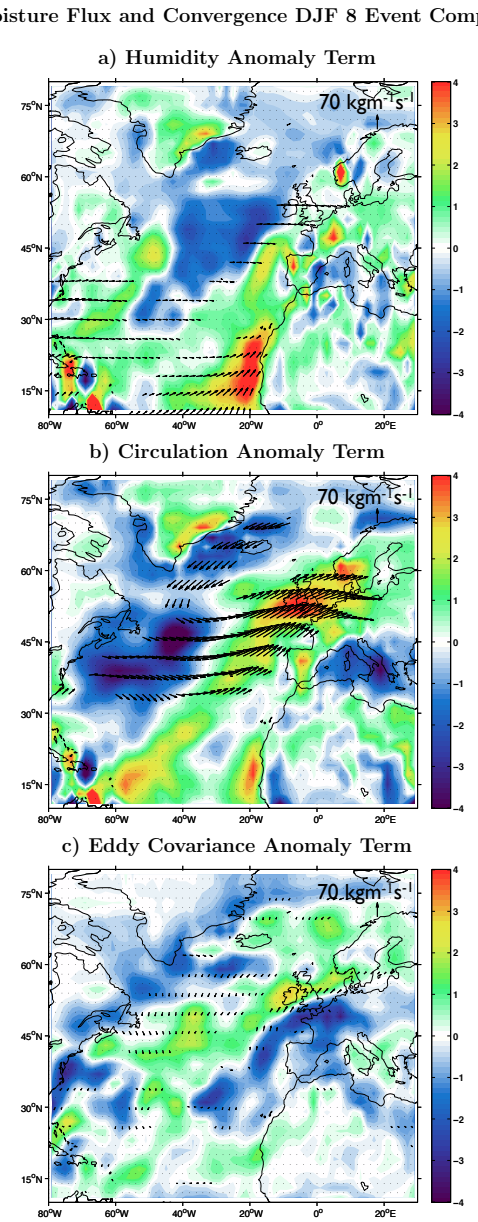


Figure 17.16: Daily decomposition of the anomalous moisture transport for the 3 Feb 2004 flood event on the River Eden at Temple Sowerby. Moisture flux in kg m^{-1} s^{-1} (strongest 20 percent of values shown as arrows) and moisture convergence in mm day^{-1} (colors), (top) humidity anomaly term, (middle) circulation anomaly term, and (bottom) eddy covariance anomaly term. (a) Day -2: 1 Feb 2004 and (b) Day -3: 31 Jan 2004.

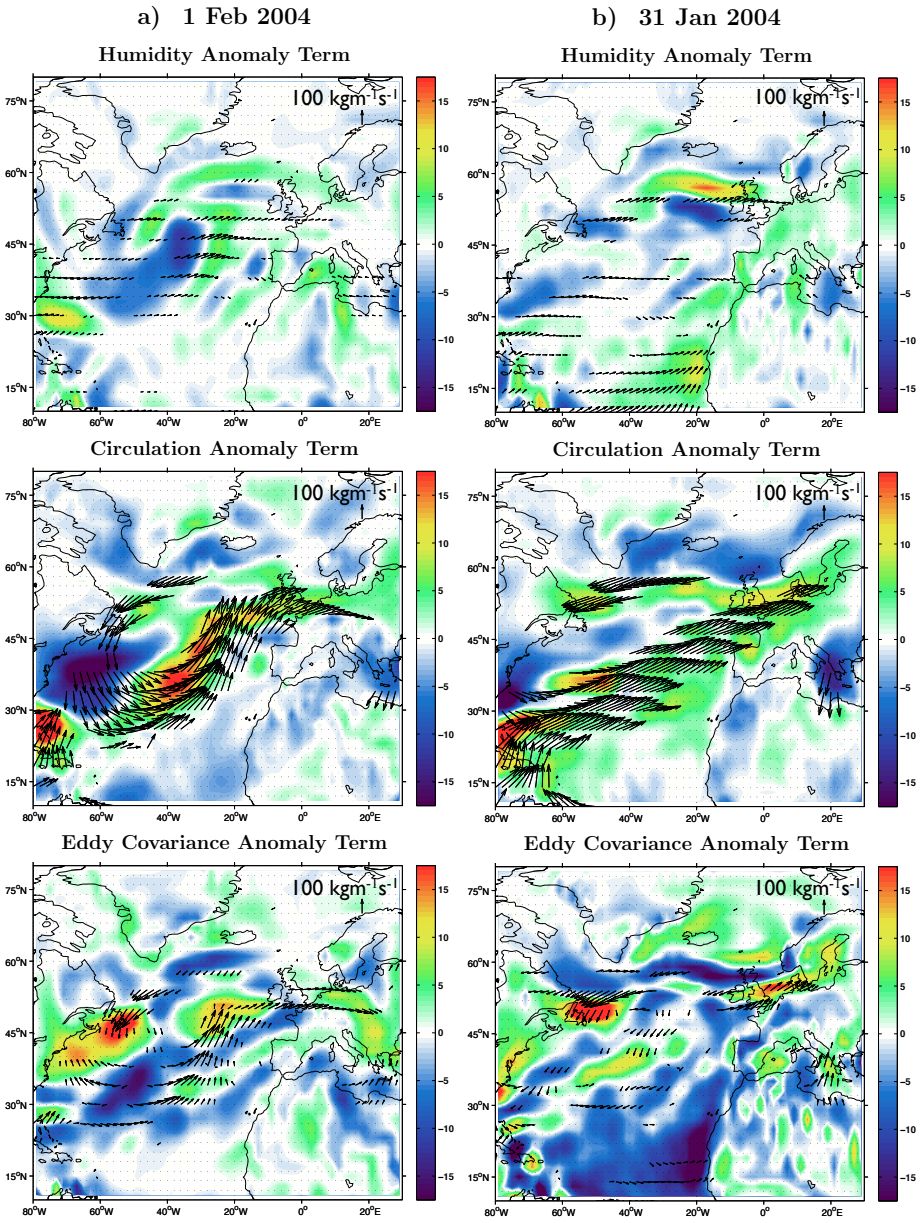


Figure 17.17: 20th Century Reanalysis V2 data (20CR, Compo et al. 2011) 900 mb specific humidity fields at 0900 UTC for one day prior to four of the 20 historical Ohio basin flood dates.

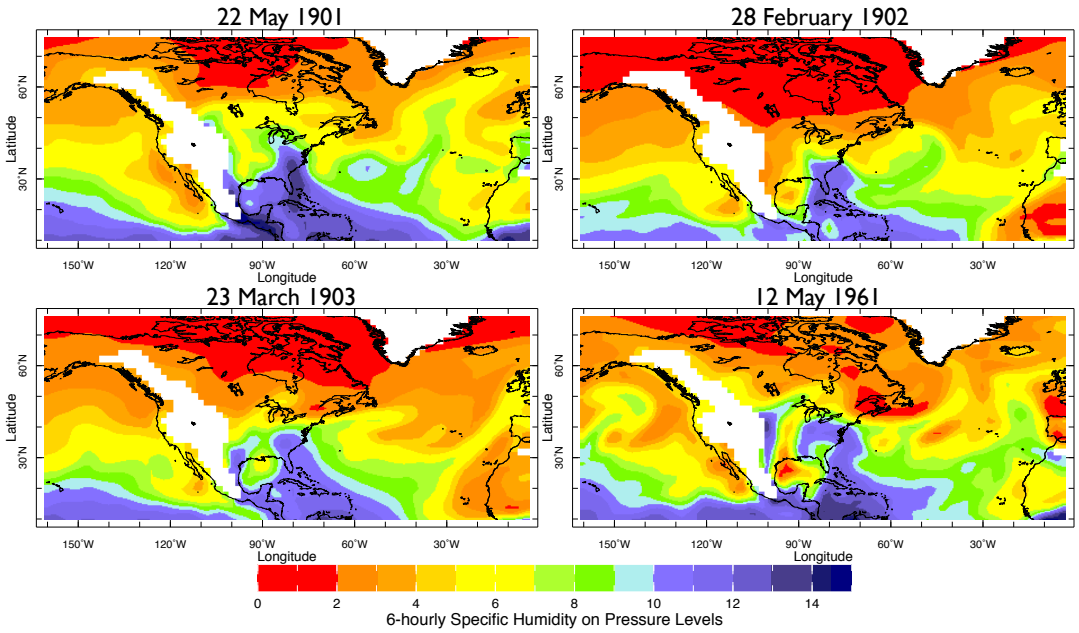


Figure 17.18: Same as Fig. 17.17, but for two days prior.

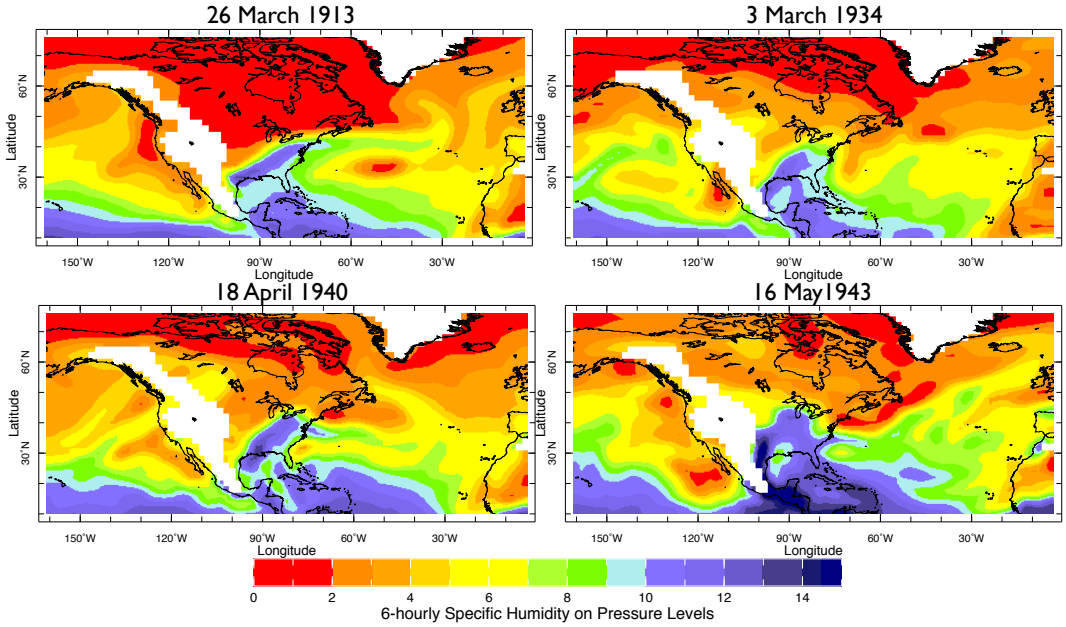


Table 17.1: Date 10-year flood threshold is exceeded at gauge station and USGS station identifier.

21 Apr 1901	03193000
23 May 1901	03193000
1 March 1902	03193000
24 March 1903	03193000
28 March 1913	03193000
20 March 1933	03234500, 03377500
14 May 1933	03234500
5 March 1934	03193000
17 April 1939	03234500
20 April 1940	03234500
20 March 1943	03234500
18 May 1943	03335500, 03360500, 03377500
7 March 1945	03234500, 03373500
22 March 1945	03234500
14 April 1948	03234500
13 May 1961	03360500, 03373500, 03374000, 03377500
5 March 1963	03234500
10 March 1964	03234500
25 May 1968	03234500
4 May 1996	03373500

Table 17.2: Attributes of the seven selected stations in the Ohio Basin.

IWMO	**RIVER**	**CITY**	**STATE**	**LON °W**	**LAT °N**	**BASIN km^2**	**Years**
03193000	Kanawha	Kanawha Falls	WVA	81.13	38.08	21,681	131
03234500	Scioto	Higby	OHIO	82.52	39.12	13,289	78
03335500	Wabash	Lafayette	IND	86.53	40.25	18,821	85
03360500	White	Newberry	IND	87.00	38.55	12,142	80
03373500	White	Shoals	IND	86.47	38.40	12,761	85
03374000	White	Petersburg	IND	87.17	38.30	28,814	80
03377500	Wabash	Mt. Carmel	ILL	87.45	38.24	74,164	69

This page intentionally left blank

Chapter 18

Conclusions

18.1 General contributions and key concepts

Parts I, II, and III are examples of the moisture transports from the N. Atlantic or Gulf of Mexico that cause extreme flooding in large basins of the Eastern or Central U.S. as illustrated by Fig 1.1. In Part I, common track groupings, differing lengths of North Atlantic tropical cyclone track information of latitude (Y) and longitude (X) every six hours were converted into a single length of standardized moment information. The first moment, the mean, provided two variables for mean centroid position (X, Y). The second moment, variance, provided three variables forming an ellipse (X, Y, XY). The five variables represent the entire track length and shape, and were normalized to weight the mean and variance equally. With the five weighted variables K-means clustering was possible and six clusters of North Atlantic tracks were found with distinct genesis locations, track shapes, intensities, life spans, landfalls, seasonality, and trends. The key concepts of this section are the groupings of tracks with common features and the method of moments that was used. Paliwal and Patwardhan (2013) adopted this method to identify clusters in the North Indian Ocean. They also used regression mixture model

and hierarchal clustering and arrived at the same clusters as the moments method, strengthening their conclusions.

In Part II, nonparametric stochastic model is created and tested for the simulation of North Atlantic tropical cyclone tracks. The model simulates tracks by randomly sampling track segments of varying length, selected from historical tracks, jumping to tracks close in distance, direction, and age. Final length of the simulated track is determined by a random draw near the genesis location. The selection of the probability criteria to jump to another track and track length were selected to statistically capture and display the dynamical behavior differences of tropical cyclones born in different parts of the basin as in Part 1. The key concept of Part II is a functional track model that can be run many ways. Part II presented two ways: as a model to reproduce many iterations of the entire historical record, and to assess probabilities of storm movement after passing through a given point in the basin. In this way, risk probabilities can be tailored to a specific region.

In Part III, dynamics and blocking, twenty 10-yr flooding events in the Ohio basin were composited and produced a coherent signal of a semi-permanent anomalous high centered at 60°W 36°N that shifted the climatological flow from the Gulf toward the basin. The anomalies of vertically integrated (600 mb to surface) moisture flux are dominated by the advection of the mean moisture field by the anomalous atmospheric circulation. A key concept of Part III is the confirmation of the Hirschboeck (1991) hypothesis that extreme floods in a given large basin and at a given season are due to typical circulation types and are not a collection of random, unrelated events. Lavers and Villarini (2013) have expanded on the work in Part III, assessing the hydrologic impact

of tropical moisture export to the central United States and further studying the atmospheric channeling known as atmospheric rivers.

18.2 Future work

The Lagrangian view of following a single fluid parcel as it moves through time and space gives tracks that reflect the dynamic organization of the system at different scales. There are several areas in climate science where flows are organized into tracks. In the atmosphere, tropical and extratropical cyclone tracks, tropical moisture exports and their concentration into atmospheric rivers, and in water with tracer paths, and buoy tracks in ocean currents. Hamilton (2013) state that "Nakamura et al. (2009) technique appears likely to provide reasonable results for buoy trajectories". These tracks are possible applications for the moments method of Part I as a question of how best to automatically identify their organization and the physical variables that control the emergence and evolution of members of each typology (cluster).

Booth et al. (2014) studied three typical paths of winter (DJF) extratropical cyclones in ERA-Interim data using feature identification of Hodges (1999) that cause high winds in the northeast U.S.: from the WNW (Alberta Clippers), from the south (Nor'easter), and from the SW; and one path that does not, from the SW and out to sea. Comparing these to the moments clustering of the same data resulted in five optimal clusters shown in Fig. 18.1, with the mean centroid (x) and variance ellipse of the cluster in bold. The mean variance ellipse of cluster 4 shows the shorter length west to east Alberta Clippers, and the longer length horizontal tracks appeared in cluster 2. Shorter SW and Nor'easter tracks appear in cluster 1, longer SW in cluster 3 and 2, and out to sea tracks in cluster 5.

Booth et al. (2014) classification is based on direction, and while the moments clustering includes direction it also includes track length and geographical region. This can be adjusted by tuning the weight of the centroid and variance portions (default is equal weight), and by adding meteorological conditions as a weighting factor. Central pressure is related to wind speed and could be added as a weight. Future work can examine criteria by which we can judge that one clustering scheme is better than another. Possible criteria are the misclassification rate, such as Alberta Clippers vs. Nor'easters, or predictability or ability to associate with what we consider driving variables from a theoretical perspective like central pressure or wind speed, or empirical predictability such as the average wind speed of each cluster and cluster frequency of occurrence. This empirical predictability is explored further with the HITS model of Part II.

A focus on meteorological conditions and climate state can be applied both to the clusters of Part I and as an increase or decrease in the number of storms in a season or as a weight to the transition probabilities in Part II to improve risk assessment applications. By compositing fields of cyclone days for each cluster to find large-scale patterns of atmospheric circulation and sea surface temperature this could be used as a test to find variables important for HITS.

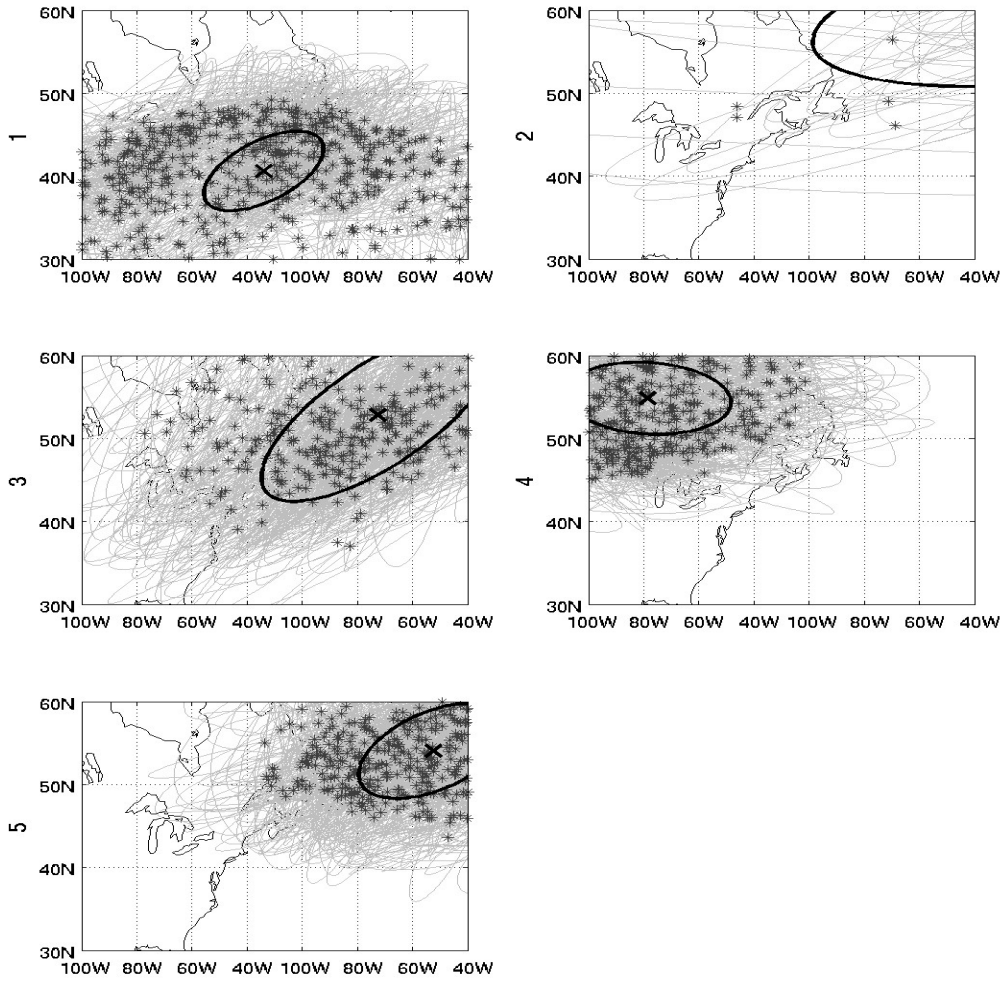
Likely candidates for future work are ENSO state and relative sea surface temperatures. El Nino has been shown to suppress North Atlantic tropical cyclone activity both by an increase in shear (Gray 1984), and an increase in tropospheric temperature that decreases the potential intensity (Tang and Neelin 2004). Relative sea surface temperature, or a measure of local sea surface temperature to the mean of the sea surface temperature of the entire tropics, is a proxy for potential intensity and is shown by Ramsey and Sobel

(2011), using an idealized model, to be more important for North Atlantic tropical cyclone behavior than absolute sea surface temperature alone.

HITS in Part II was run in chapter 10.5 to assess probabilities of storm movement after passing through a given point, but could also be adapted to run backwards. In this way the user would be able to see storm genesis regions and tracks of tropical cyclones that ended in a particular box. Combining this with the climate information could tailor a probability for a region and climate state.

Research questions relating to Part III can be a generalized to both floods and droughts. There are strong blocking patterns that translate into persistent rain leading to floods and also to the maintenance of drought. These persistent blocks, typically with low wave numbers, are not well reproduced by numerical models (Anderson 1993). This leads to questions of how such patterns and their properties can be better identified from data and related to model state variables. The dipole index of chapter 15.2 was a simple index of blocking, this can be expanded by looking at the whole hemispheric pattern with EOFs as in Lu et. al (2013). The international community has produced historical and model projections at an increased resolution (CMIP5), which may provide further insights. A smaller resolution gives smaller finite differences in the partial derivatives of the equations of motion leading to less numerical diffusion. It is possible that less diffusion would allow blocks in the model to persist.

Figure 18.1: Moment clustering of winter (DJF) extratropical cylone tracks in ERA-Interim data. Centroid locations (stars) and directional variance (light ellipses) for the five clusters. Mean centroid value is marked with a dark bold x, and mean variance ellipse with a dark bold line.



This page intentionally left blank

Part V

Bibliography

This page intentionally left blank

Bibliography

Anderson, J. L., 1993: The climatology of blocking in a numerical forecast model. *Journal of Climate*, **6**(6), 1041-1056.

Avila, L.A., 2013: Tropical Cyclone Report, Hurricane Rafael (AL172012), 12-17 October 2012, *National Hurricane Center*. United States National Oceanic and Atmospheric Administration's National Weather Service.

Barriopedro, D., E. M. Fischer, J. Luterbacher, R. M. Trigo, and R. García-Herrera, 2011: The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe, *Science*, **332**, 220-224.

Berg, R.J., 2013: Tropical Cyclone Report, Hurricane Isaac (AL092012), 21 August - 1 September 2012, *National Hurricane Center*. United States National Oceanic and Atmospheric Administration's National Weather Service.

Bhat, U. N., & Miller, G. K. (1972). *Elements of applied stochastic processes*. J. Wiley.

Blake, E.S., Kimberlain, T.B., Berg, R.J., Cangialosi, J.P., and Beven II, J.L., 2013: Tropical Cyclone Report, Hurricane Sandy (AL182012), 22-29 October 2012, *National Hurricane Center*. United States National Oceanic and Atmospheric Administration's National Weather Service.

Blender, R., K. Fraedrich, and F. Lunkeit, 1997: Identification of cyclone-track regimes in the North Atlantic. *Quart. J. Roy. Meteor. Soc.*, **123**, 727-741.

Booth, J.F., Rieder, H.E., Lee, D.E., and Y. Kushnir, 2014: Unraveling the characteristic paths of extratropical cyclones that cause high wind events in the Northeast United States. *JGR*, submitted.

Buchman, S.M., Lee, A.B., and Schafer, C.M., 2011: High-dimensional density estimation via SCA: An example of modelling of hurricane tracks, *Statistical Methodology*, **8**(1), pp. 18-30.

Camargo, S.J. and J.P. Kossin, 2008. Relationship of the Atlantic Meridional Mode with Atlantic hurricane tracks. 28th Conference on Hurricanes and Tropical Meteorology, P2.A2. 28 April - 2 May 2008, Orlando, FL. American Meteorological Society, Boston.

Camargo, S.J., A.W. Robertson, S.J. Gaffney, P. Smyth, and M. Ghil, 2007: Cluster Analysis of Tropical Cyclone Tracks. Part I: General Properties. *J. Clim*., **20**, 3635-3653.

Casson, E. and Coles, S., 2000: Simulation and extremal analysis of hurricane events, *Appl. Statist.*, **49**(Part 2), pp. 227-245.

Chu, P-S and Wang, J., 1998: Modeling Return Periods of Tropical Cyclone Intensities in the Vicinity of Hawaii, *J. Appl. Meteorol*., **37**, pp. 951-960.

Çinlar, E., 1975: Exceptional Paper—Markov Renewal Theory: A Survey.*Management Science*, *21*(7), 727-752.

Cinlar, E., 1969: Markov renewal theory. *Advances in Applied Probability*, *1*(2), 123-187.

Clark, K.M., 1986: A formal approach to catastrophe risk assessment and management, *Proceedings (Casualty Actuarial Society)*, **73**(Part 2), pp. 69-92.

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, N. Matsui, R.J. Allan, X. Yin, B.E. Gleason, R.S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R.I. Crouthamel, A.N. Grant, P.Y. Groisman, P.D. Jones, M. Kruk, A.C. Kruger, G.J. Marshall, M. Maugeri, H.Y. Mok, Ø. Nordli, T.F. Ross, R.M. Trigo, X.L. Wang, S.D. Woodruff, and S.J. Worley, 2011: The Twentieth Century Reanalysis. *Quarterly J. Roy. Meteorol. Soc.*, **137**, 1-28.

Dettinger, M. (2011), Climate change, atmospheric rivers, and floods in California—A multimodel analysis of storm frequency and magnitude changes, *J. Am. Water Resour. Assoc., 47(3),* 514–523.

Dettinger, M.D., Ralph, F.M., Das, T., Neiman, P.J., and Cayan, D., 2011: Atmospheric rivers, floods, and the water resources of California. *Water*, **3** (Special Issue on Managing Water Resources and Development in a Changing Climate), 455-478.

Dirmeyer, P. A. and J. L. Kinter III, 2009: 'The "Maya Express": Floods in the U.S. Midwest', *Eos Trans. AGU*, **90**(12), 101.

Dirmeyer, P. A. and J. L. Kinter III, 2010: Floods over the U.S. Midwest: A Regional Water Cycle Perspective. J. Hydrometeor, **11**, 1172-1181.
Dole, R. et al., 2011: Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.*, **38**, L06702.

Elsner, J.B., 2003: Tracking hurricanes. *Bull. Amer. Meteor. Soc.,* **84**, 353-356.

Elsner, J.B. and Kara, A.B., 1999: *Hurricanes of the North Atlantic: Clim. and Society*, Oxford University Press, pp. 1-512.

Elsner, J.B., J.P. Kossin, and T.H. Jagger, 2008: The increasing intensity of the strongest tropical cyclones. *Nature*, **455**, 92-95.

Elsner, J.B. and K. B. Liu, 2003: Examining the ENSO-typhoon hypothesis. *Clim. Res.,* **25**, 43-54.

Emanuel, K.A., 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686-688.

Emanuel, K., 2007: Environmental Factors Affecting Tropical Cyclone Power Dissipation. *J. Clim.*, **20**, 5497-5509.

Emanuel, K., 2010: Tropical Cyclone Activity Downscaled from NOAA-CIRES Reanalysis, 1908-1958. *J. Adv. Model. Earth Syst.*, **2**, Art. #1, 12 pp.

Emanuel, K. and Jagger, T., 2010: On Estimating Return Periods. *J. Appl. Meteorol. Clim.,* **49**, pp. 837-844.

Emanuel, K., Ravela, S., Vivant, E. and C. Risi. 2006: A Statistical-Deterministic Approach to Hurricane Risk Assessment. *Bull. Amer. Meteor. Soc.,* **87**, pp. 299–314.

Federal Interagency Floodplain Management Task Force (FIFMTF), 1992: "Floodplain Management in the United States: An Assessment Report," Volume 1, Summary Report. Washington, DC: Federal Emergency Management Agency, 69 pp.

Foufoula‐Georgiou, E., & Lettenmaier, D. P. (1987). A Markov renewal model for rainfall occurrences. *Water Resources Research*, *23*(5), 875-884.

Gilbert, G., Peterson, G. L., & Schofer, J. L. (1972). Markov renewal model of linked trip travel behavior. *Journal of the Transportation Engineering Division*,*98*(3), 691-704.

Goldenberg, S. B., C. W. Landsea, A. M. Mestas-Nunez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications. *Science*, **293,** 474-479.

Gray, W.M., 1968: Global View of the Origin of Tropical Disturbances and Storms. *Mon. Wea. Rev*., **96**, 669-697.

Gray, W.M., 1979: Hurricanes: Their formation, structure and likely role in the tropical circulation. *Supplement to: Meteorology Over the Tropical Oceans*, D. B. Shaw, Ed., Royal Meteorological Society, 155-218.

Gray, W. M., 1984: Atlantic seasonal hurricane frequency. Part I: El Nino and 30 mb quasi-biennial oscillation influences, *Mon. Weather Rev.*, **112**, 1649– 1668.

Halevy, A., Norvig, P., & Pereira, F., 2009: The unreasonable effectiveness of data. *Intelligent Systems, IEEE*, *24*(2), 8-12.

Hall, T.M. and Jewson, S., 2007: Statistical modelling of North Atlantic tropical cyclone tracks. *Tellus*, **59A**, pp. 486-498.

Hall, T.M. and Jewson, S., 2008: SST and North American Tropical Cyclone Landfall: A Statistical Modeling Study. arXiv:0801.1013v1[physics.ao-ph].

Hallegate, S., 2007: The Use of Synthetic Hurricane Tracks in Risk Analysis and Climate Change Damage Assessment. *J. Appl. Meteorol. Clim.*, **46**, pp. 1956-1966.

Hamiton, L.J., 2013: Statistical clustering of drifting buoy trajectories to identify Lagrangian circulation features around Japan and off Fukushima. *Methods in Oceanography*, **6**, 16-32.

Henderson-Sellers, A., H. Zhang, G. Berz, K. Emanuel, W. Gray, C. Landsea, G. Holland, J. Lighthill, S. L. Shieh, P. Webster, and K. McGuffie, 1998: Tropical cyclones and global climate change: A post-IPCC assessment. *Bull. Amer. Meteorol. Soc.*, **79,** 19-38.

Higgins, R. W., K. C. Mo, 1997: Persistent North Pacific Circulation Anomalies and the Tropical Intraseasonal Oscillation. *J. Climate*, **10**, 223–244.

Hirschboeck, K. K., 1991: National Water Summary 1988-89 Floods and Droughts: Hydrologic Perspectives on Water Issues.

Hodges, K. I., 1999: Adaptive Constraints for Feature Tracking. *Mon. Wea. Rev*., **127**, 1362–1373.

Hoskins, B. J., F. P. Bretherton, 1972: Atmospheric Frontogenesis Models: Mathematical Formulation and Solution. *J. Atmos. Sci.*, **29**, 11–37.

Inwards, R., 1898: *Weather Lore*. Elliot Stock, London, p. 86.

Jarvinen, B. R., C. J. Neumann, and M. A. S. Davis, 1984: A tropical cyclone data tape for the North Atlantic Basin, 1886-1983: Contents, limitations, and uses. *NOAA Technical Memorandum* NHC 22, Coral Gables, Florida, pp. 1-21.

Katz, Richard W., 2002: Stochastic Modeling of Hurricane Damage. *J. Appl. Meteor.*, **41**, 754–762.

Kaufman, L. and P. J. Rousseeuw, 1990: *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, Wiley Series in Probability and Mathematical Statistics.

Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fiorino, 2001: The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247-268.

Knippertz, P., and H. Wernli, 2010: A Lagrangian Climatology of Tropical Moisture Exports to the Northern Hemispheric Extratropics. *J. Climate*, **23**, 987-1003.

Knippertz, P., Wernli, H., and G. Gläser, 2013: A Global Climatology of Tropical Moisture Exports. *J. Climate*, **26**, 3031–3045.

Koenker, R. and G. Bassett Jr., 1978: Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50.

Kossin, J.P., K.R. Knapp, D.J. Vimont, R.J. Murnane, and B.A. Harper, 2007: A globally consistent reanalysis of hurricane variability and trends. *Geophys. Res. Lett.*, **34**, L0815.

Kushnir, Y., and J. M. Wallace, 1989: Low-Frequency Variability in the Northern Hemisphere Winter - Geographical-Distribution, Structure and Time-Scale Dependence, *J. Atmos. Sci.*, **46**(20), 3122-3142.

Landsea, 2005: Hurricanes and Global Warming. *Nature*, **438**, E11-E13.

Lavers, D. A., Allan, R.P., Wood, E.F., Villarini, G., Brayshaw, D.J. and A. J. Wade, 2011: Winter floods in Britain are connected to atmospheric rivers, *Geophys. Res. Lett.*, **38**, L23803.

Lavers, D.A and G. Villarini, 2013: Atmospheric Rivers and Flooding over the Central United States, *J. Clim.*, **26**, 7829-7836.

Li, Wenhong, Laifang Li, Rong Fu, Yi Deng, Hui Wang, 2011: Changes to the North Atlantic Subtropical High and Its Role in the Intensification of Summer Rainfall Variability in the Southeastern United States. *J. Climate*, **24**, 1499–1506.

Loader, C., 1999: *Local Regression and Likelihood*. Springer, 309pp.

Lott, F., A. W. Robertson, M. Ghil, 2004: Mountain Torques and Northern Hemisphere Low-Frequency Variability. Part II: Regional Aspects. *J. Atmos. Sci.*, **61**, 1272–1283.

Lu, M., U. Lall, A. Schwartz, and H. Kwon, 2013: Precipitation predictability associated with tropical moisture exports and circulation patterns for a major flood in France in 1995, *Water Resour. Res.*, **49**, 6381–6392.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. *Proc. Fifth Berkeley Symp. On Mathematical Statistics and Probability,* Berkeley, CA, University of California, 281-297.

Milrad, Shawn M., Eyad H. Atallah, J.R. Gyakum, 2010: Synoptic Typing of Extreme Cool-Season Precipitation Events at St. John's, Newfoundland, 1979–2005. *Wea. Forecasting*, **25**, 562–586.

Moore, B.J., P.J. Neiman, F.M. Ralph, and F.E. Barthold, 2012: Physical Processes Associated with Heavy Flooding Rainfall in Nashville, Tennessee, and Vicinity during 1-2 May 2010: The Role of an Atmospheric River and Mesoscale Convective Systems, *Mon. Wea. Rev.*, **140**, 358-378.

Nakamura, J., U. Lall, Y. Kushnir and S. J. Camargo, 2009: Classifying North Atlantic Tropical Cyclone Tracks by Mass Moments. *Journal of Climate*, **22**(20), 5481-5494.

Neiman, P.J., L.J. Schick, F.M. Ralph, M. Hughes, and G.A. Wick, 2011: Flooding in Western Washington: The Connection to Atmospheric Rivers. *J. Hydrometeo.*, **12**, 1337-1358.

Neumann, C.J., Jarvinen, B.R., McAdie, C.J., and Elms, J.D. 1993: *Tropical Cyclones of the North Atlantic Ocean 1871-1992*. National Climate Data Center in cooperation with the National Hurricane Center, Coral Gables, FL, 193pp.

Ockenden, M.C. and N.A. Chappell, 2011: Identification of the dominate runoff pathways from the data-based mechanistic modelling of nested catchments in the temperate UK, *J. Hydro.*, **402**, 71-79.

Paliwal, M. and A. Patwardhan, 2013: Identification of clusters in tropical cyclone tracks of North Indian Ocean, *Nat. Hazard*, **68**, 645-656.

Perry, C. A., 2000: Significant floods in the United States during the 20th century - USGS measures a century of floods. USGS Fact Sheet, 024-00.

Pielke, Jr. R. A., and Landsea, C. W. 1998: Normalized Atlantic hurricane damage, 1925-1995. *Wea. Forecasting*, **13**, 621-631.Katz, R. W., M. B. Parlange, and C. Tebaldi, 2003: Stochastic modeling of the effects of large-scale circulation on daily weather in the southeastern U.S. *Climatic Change*, **60**, 189-216.

Ralph, F.M., and M.D. Dettinger, 2011: Storms, Floods, and the Science of Atmospheric Rivers. *EOS*, **92**(32), 265-266.

Ralph, F. M., P. J. Neiman, G.A. Wick, S.I. Gutman, M.D. Dettinger, D.R. Cayan, and A.B. White, 2006: Flooding on California's Russian River: Role of atmospheric rivers. *Geophys. Res. Lett.*, **33**, L13801.

Ramsay, H.A., and A.H. Sobel, 2011: The effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single column model. *J. Climate*, **24**, 183-193.

Robertson, A. W., Y. Kushnir, U. Lall, and J. Nakamura, 2012: Weather and Climatic Drivers of Extreme Flooding Events over the Midwest of the United States. In: AGU Monograph "Extreme Events: Observations, Modeling and Economics", M. Chavez, M. Ghil, J. Urrutia-Fucugauchi (Eds.), submitted.

Rumpf J., Weindl H., Höppe P., Rauch E., and Schmidt V., 2007: Stochastic modelling of tropical cyclone tracks. *Math. Meth. Oper. Res*., **66**(3), pp. 475–490.

Rumpf J., Weindl H., Höppe P., Rauch E., and Schmidt V., 2009: Tropical cyclone hazard assessment using model-based track simulation. *Nat. Hazards*, **48**, pp. 383-398.

Shepherd, J.M., Grundstein, A., and T.L. Mote, 2007: Quantifying the contribution of tropical cyclones to extreme rainfall along the coastal southeastern United States. *GRL*, **34**(23), L23810, doi:10.1029/2007GL031694.

Smith, J. A., Baeck, M. L., Villarini, G., & Krajewski, W. F., 2010: The Hydrology and Hydrometeorology of Flooding in the Delaware River Basin. *Journal of Hydrometeorology*, **11**(4), pp. 841-859.

Tang, B. H. and J. D. Neelin, 2004: ENSO Influence on Atlantic Hurricanes via Tropospheric Warming. *Geophys. Res. Lett.*, **31**, L24204, doi:10.1029/2004GL021072.

Truchelut, R. E. and R. E. Hart, 2011: Quantifying the possible existence of undocumented Atlantic warm-core cyclones in NOAA/CIRES 20th Century Reanalysis data. *Geophys. Res. Lett.*, **38**, L08811.

Vickery, P.J., Skerlj, P.F., Twisdale Jr., L.A., 2000: Simulation of hurricane risk in the US using an empirical track model. *J. Struct. Eng.*, **126**(10), pp. 1222–1237.

Vose, R. S., Richard L. Schmoyer, Peter M. Steurer, Thomas C. Peterson, Richard Heim, Thomas R. Karl, and J. Eischeid, 1992: The Global Historical Climatology Network: long-term monthly temperature, precipitation, sea level pressure, and station pressure data. ORNL/CDIAC-53, NDP-041. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Yonekura, E., and T.M. Hall, 2011: A statistical model of tropical cyclone tracks in the Western North Pacific with ENSO-dependent cyclogenesis. *J. Appl. Meteorol. Climatol.*, **50**, pp. 1725-1739.

Zhu, Y. and R. E. Newell, 1994: Atmospheric Rivers and Bombs. *Geophys. Res. Lett.*, **21**, 1999-2002.

Zhu, Y. and R. E. Newell, 1998: A Proposed Algorithm for Moisture Fluxes from Atmospheric Rivers. *Mon. Wea. Rev.*, **126**, 725-735.

This page intentionally left blank